WIRELESS COMMUNICATIONS AND NETWORKS

3G and Beyond

Second Edition

ABOUT THE AUTHOR

Iti Saha Misra is currently Professor in the Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata.

She completed her BTech degree in Radio Physics and Electronics from Calcutta University (1989) and Masters degree in Telecommunication Engineering from Jadavpur University (1991), Kolkata. After the completion of her PhD degree in Engineering in the field of Microstrip Antennas from Jadavpur University, she has actively engaged herself in teaching since 1997. Her current research interests are in the areas of Cognitive Radio Networks, Call Admission Control in Cellular and WiMAX Networks, Packet Scheduling for Broadband Communication Networks, Channel Allocation and Power Management in Macro/Femto Networks, Mobility Management, Network Architecture and Protocols, Integration Architecture of WLAN and 3G Networks, Location Management for Cellular Wireless Networks. Her other research activities are related to Design Optimization of Wire Antennas using Numerical Techniques.

She has authored more than 140 journals and international conference research papers. She has presented research papers in the IEEE and other reputed international conferences, delivered invited lectures, conducted sessions as Session Chair in the USA, UK, France, Australia, Prague, Singapore, Malaysia, Thailand and Bangladesh. She was the recipient of the Career Award for Young Teachers (CAYT) by the All India Council for Technical Education (AICTE) in the financial year 2003–2004 and received the IETE Gowri Memorial Award for presenting the best paper in the general topic of *4G Networks: Migration to the Future.*

One of her papers titled "Design and study of VoIP Model in Cognitive Radio Network under Different Simulation Platforms," received the best paper award in CUBE, ACM International Information Technology Conference. She has been directly associated with the development of a Broadband Wireless Communication Laboratory in the Department of Electronics and Telecommunication Engineering, Jadavpur University under the DST FIST project, and has also implemented Voice over Wireless LAN for the advanced research in this domain.

She is a senior member of IEEE, founder chair of Women in Engineering, Affinity Group, IEEE Calcutta Section and is presently Secretary of IEEE Kolkata Section.

WIRELESS COMMUNICATIONS AND NETWORKS

3G and Beyond

Second Edition

Iti Saha Misra
Professor
Department of Electronics and Telecommunication Engineering
Jadavpur University
Kolkata, West Bengal

McGraw Hill Education (India) Private Limited
NEW DELHI

McGraw Hill Education Offices

New Delhi New York St Louis San Francisco Auckland Bogotá Caracas Kuala Lumpur Lisbon London Madrid Mexico City Milan Montreal San Juan Santiago Singapore Sydney Tokyo Toronto

McGraw Hill Education (India) Private Limited

Published by McGraw Hill Education (India) Private Limited,
P-24, Green Park Extension, New Delhi 110 016.

Wireless Communications and Networks: 3G and Beyond, 2e

ISBN (13): 978-1-25-906273-5
ISBN (10): 1-25-906273-2

Vice President and Managing Director: *Ajay Shukla*

Head—Higher Education Publishing and Marketing: *Vibha Mahajan*
Publishing Manager—SEM & Tech Ed.: *Shalini Jha*
Editorial Executive: *Koyel Ghosh*
Manager—Production Systems: *Satinder S Baveja*
Copy Editor: *Preyoshi Kundu*
Sr Production Executive: *Suhaib Ali*

Asst General Manager—Higher Education Marketing: *Vijay Sarathi*
Sr Product Specialist: *Tina Jajoriya*
Sr Graphic Designer—Cover: *Meenu Raghav*

General Manager—Production: *Rajender P Ghansela*
Production Manager: *Reji Kumar*

Typeset at BeSpoke Integrated Solutions, Puducherry 605 008, India. Printed at Magic International Pvt. Ltd., Plot No. 26 E, Site IV (Industrial), Sector-31, Greater Noida

Cover Printer: Magic International Pvt. Ltd.

RXCLCRQODQLLR

CONTENTS

PREFACE

Telecommunication has been the most widely discussed topic over the past few decades – from cell phone to smartphone, Wi-Fi to WiMAX, wireless communication is becoming one of the fastest growing field of technology. We are passing through the third phase of growth with the expected video traffic communications. The first phase was for voice only communication, and the second phase was for voice communication with limited data services. Portable devices, cellular networks, Wireless LAN, Broadband wireless access of 3G LTE, 3G LTE-A, WiMax Networks were the main enablers that fulfilled our dream in wireless communications. We cannot live without the Internet, as information exchange has become an important aspect of our lives. The second generation GSM does not have much capability to deliver value-added Internet traffic, and thus a huge involvement in research and investment of resources is being made for the development of 3G and beyond technology. To keep pace with the system and application requirements, Broadband Wireless Communication (BWC) is becoming a cutting edge technology for no limits communication. The possibility for any user, from any place and at any time, to carry out any type of transmission whether it is voice, data, images or video, is causing a revolution in the world of communications. Wireless Local Area Network, Personal Area Network and Broadband Wireless Access, etc., have flooded countries worldwide.

The first edition of *Wireless Communication and Networks* was published in 2009 and was well received by teachers and students all over the country. The concepts of this book present the fundamentals of wireless communications and networks, evolution, functionalities in a simple way for self-learning remain same as in the previous edition, with the addition of four new chapters. This was essential in order to provide up-to-date information to students.

New to this Edition

New chapters are based on the following topics:

- Digital Modulation fundamentals covering MSK, GMSK, OFDM modulation techniques
- Equalization, Diversity and Coding for fading channels covering Probability of Error, BER, etc.
- Cellular Wireless Networks with the in-depth coverage of design fundamentals, communication principles, handoff process etc.
- Characteristics of Wireless Channels and Propagation Path Loss Models covering the effect of multipath propagation and time variant channel characterization
- Appendices on 3G LTE and Bluetooth technology

Salient Features of the Book

- Digital modulation and demodulation techniques adopted to help understand the evolution of wireless communication technologies
- Extensive discussions on propagation models
- Focus on cellular concepts to build on latest wireless technology
- Includes the worldwide revolution of latest mobile technologies – Wi-MAX (Worldwide Interoperability for Microwave Access) and Wi-Fi (Wireless Fidelity)
- Coverage of Universal Mobile Telecommunication Systems (UMTS)—The 3G Networks
- Detailed discussion on Mobile IP with its limitations and solution strategy
- Fundamental and theoretical concepts are presented lucidly with sufficient illustrations and explanations.

- Numerous solved examples and self-test exercises to enhance the understanding of the subject
- Sample questions provided to judge students' learning
- Revised pedagogy includes
  - 40 Solved examples
  - 277 Practice questions
  - 241 Objective-type questions

Organization of the Book

The present edition contains 16 chapters and 2 appendices.

The second edition includes four new chapters with sufficient in-depth study of subjects. These chapters include the fundamentals of wireless cellular communication, propagation effects and channel characteristics along with path loss models, digital modulation techniques in wireless communication, equalization, diversity and coding, the receiver techniques. In the appendix section, overviews of 3G LTE and 3G LTE-A, Bluetooth technology have been provided to touch upon all the essential topics in wireless communication.

Chapter 1 gives an introduction to Wireless Communication Networks. A wireless network is the basic platform for mobile users which provides the ability to communicate with people on the move. With the advancement of digital technology and miniaturization of circuits, the wireless communication networks have evolved from low capacity to high capacity with the support for increased data rate.

Chapter 2 discusses the evolution of modern mobile wireless communication systems. This chapter demonstrates the evolutionary path of modern wireless communication networks starting from the first Generation (1G) analog systems to fourth generation (4G) IP systems along with the development of Wireless Local Area Networks (WLAN).

Chapter 3 deals with cellular mobile wireless networks with system design and fundamentals with the addition of many new topics such as hand-off process, priority hand-off model, call admission control, variants of channel assignments schemes and the essence of radio resource management.

Chapter 4 is a completely new chapter on characteristics of wireless channels and propagation path loss models. Understanding of wireless channels is the most important aspect to grasp the subject. So, emphasis has been given to this chapter with mathematical models, analysis, numerical examples and illustrations with MATLAB output. This chapter is the essential part of studies both for undergraduate and postgraduate students. The role of antenna is major for any wireless communications system. Design of antenna for base station and mobile station are different. Propagation multipath effect further influences antenna design. At the end of this chapter, overview of mobile communication antennas for base station and mobile station are also given.

Chapter 5 is also a new chapter on digital modulations in wireless communications. The data transmission over wireless channel is digital. Today's mobile communication systems extensively use digital modulation techniques. Some of the most important digital modulation techniques are Binary Phase Shift Keying (BPSK), M-ary Phase Shift Keying (MPSK), Minimum Phase Shift Keying (MSK), Gaussian filtered MSK (GMSK), and Orthogonal Frequency Division Multiplexing (OFDM). Among these techniques, GMSK is used for GSM cellular networks and OFDM is used in WLAN and WiMAX system. The main issue of concern in any communication system is to design optimum filter at the receiver with the consideration of modulation techniques and transmission schemes such that the error probability gets minimized in the presence of noise. Issues such as calculation of bit error rate and spectrum occupancy, modulation and demodulation implementation are covered. The highlight of this chapter is the incorporation of hardware approach and results from vector signal analyzer.

Chapter 6 discusses the three independent techniques that are used separately or in combination to process the signal to improve received signal quality and link performance within small scale time and distances — equalization, diversity, and channel coding techniques. Equalization technique compensates for ISI created by multipath within time dispersive channels. Diversity techniques can be used to improve system performance in fading channels. Channel coding techniques improve the small-scale link performance by adding

redundant bits in the transmitted message and are used to detect or correct the errors in any received bit. An important aspect of this chapter is the incorporation of the details of channel response and its mathematical analysis in getting the equalization of the channel in terms of numerical examples. Illustrations are given from MATLAB program outputs for many numerical examples for both equalization and diversity techniques.

Chapter 7 deals with multiple access technologies for wireless communication systems. Sharing of limited radio resources among multiple users in any communication system is referred as *multiple access technology*. There are several multiple access technologies like frequency division multiple access (FDMA), time division multiple access (TDMA) and code division multiple access (CDMA). With the evolution of wireless networks through generations 1G to 3G, method of access technology also changes, FDMA is used for 1G, TDMA for 2G and CDMA for 3G. There are other variants of access technology for 4G communication used in 3G LTE and WiMAX networks known as Orthogonal Frequency Division Multiple Access (OFDMA).

Chapter 8 explains the Second Generation Mobile Networks-GSM: Architecture and Protocols. The first digital cellular communication is the GSM (Global System for mobile communication) released by European Telecommunications mainly for voice communication. The basic network architecture of GSM, GSM frame structure and TDMA-based access, call set-up procedure, channel types, and signaling protocols are provided to build the basic understanding of 2G cellular system.

Chapter 9 covers the 2.5G networks—the General Packet Radio Services (GPRS), which is the GSM variant, evolved both for circuit switch and packet switch communications. GPRS adds a network infrastructure based on IP protocol, which is designed with the needs of data transport. In-depth coverage of GPRS network architecture, protocols and mobility management issues have been provided. This chapter helps generate the fundamental concepts of data packet transport over cellular network.

Chapter 10 deals with the historical development of CDMA technology along with an overview of IS-95 cellular network standards. CDMA was created to provide secure communication and navigation systems for military applications. It requires the development of spread spectrum technology for multiple accesses over a single carrier frequency and reduction of interference. The advantage of CDMA for personal communication services is its ability to accommodate many users on the same frequency at the same time. The power control mechanism and hand-off process for CDMA system is described in this chapter.

Chapter 11 deals with 3G – The Universal Mobile Telecommunication System (UMTS) based on wide band code division multiple access (WCDMA) technology. UMTS is developed by ETSI (European Telecommunications Standards Institute) that belongs to the family of International Mobile Telecommunication –2000 (IMT-2000). The aim of the 3G UMTS network is to support high-speed packet data transport. The paradigm shift from the focus on voice and low speed data services to high-speed multimedia communication has generated the requirement of 3G network systems based on CDMA technology. This chapter explores the architecture and protocols of UMTS networks. The changes needed for this network are to support all-IP infrastructure.

Chapter 12 gives the overview of Internet and Mobile Internet Protocols that are needed for understanding the basics of IP transport. Mobile IP (MIP) plays an important role for wireless mobile communication. Details of MIP, from addressing to protocol and MIP variants are provided in this chapter.

Chapter 13 discusses the mobility management issues and role of IP on wireless networks, which is an important topic in this book. This chapter discusses how IP is applied in GPRS and UMTS networks and mobility is managed. Limitations of mobile IP and its advanced version MIPv6 and Hierarchical MIPv6 are also discussed in this chapter.

Chapter 14 is about the fundamentals of Wireless Local Area Networks (WLAN). It is the latest wireless access technology widely adopted in the corporate office, universities and hope for high speed Internet access. It also sometimes referred as Wireless Fidelity (Wi-Fi). The low cost access of WLAN has changed the world around us. This chapter provides the basic transmission technology, architecture, protocols, and families of WLAN. The basics of security, Quality of service (QoS) and roaming scenarios for WLAN are also discussed.

Chapter 15 discusses the Cellular and WLAN Integration—Heterogeneous Network Architecture—step towards 4G networks which is a unique feature of this book. The requirements for integration architecture, its benefits, suitable point of integrations and the types of integration are discussed at length. In today's scenario,

research in the direction of integration between the GPRS/UMTS cellular and WLAN gets importance. The primary objective of this integration is to obtain the best of the two technologies. Cellular provides wider coverage but slow data rate, whereas WLAN provides higher data rate with limited coverage. The integration between any heterogeneous access technologies creates challenges for vertical handoff and mobility management. This chapter explores the different current integration techniques and mobility management procedures.

Chapter 16 provides the detailed overview of broadband wireless technology—a revolutionary cutting edge technology–that brought the wireless and Internet revolution to portable devices across the globe. It is better acclaimed as the Worldwide Interoperability for Microwave Access (WiMaX). The IEEE 802.16 family of standards and its associated industry consortium, WiMax, promises to deliver high data rates over large areas to a large number of users in the near future. This WiMaX addition to current broadband options such as DSL (Digital Subscriber Line), cable, and WiFi promises to rapidly provide broadband access the world's rural and developing areas where broadband is currently unavailable.

The two appendices at the end give an overview of two important wireless technologies.

Appendix A discusses the most discussed topic nowadays and the evolved emerging platform for 4G communication–3G LTE and 3G LTE-A. The successful evolution and deployment of GSM family of technologies, generally known as 3GPP family have gone through the development phase of GSM, EDGE, UMTS, HSPA, HAPA+, LTE and LTE-A. In the commercial market, HSPA+ continues its progress and in the way LTE revolution began. Long Term Evolution (LTE) is the next step forward in cellular 3G services.

Appendix B deals with another very important wireless communication aspect—the Bluetooth technology. Bluetooth is the short-range radio link technology developed with the intention to replace the cable connecting portable electronic devices. Bluetooth radio modules operate in the unlicensed ISM band at 2.4GHz, and avoid interference from other signals by hopping to a new frequency after transmitting or receiving a packet. The beauty of this technology is its robustness, low complexity, low power, and low cost and its design which helps to operate in noisy environments.

Online Learning Center

- **For Instructors**
  Solution Manual, PowerPoint Lecture Slide
- **For Students**
  Interactive quiz, Model question paper with solution

Acknowledgements

The painstaking job of writing a book as a sole author is very difficult and has taken many sleepless nights. I am lucky to have such a loving family; my sincere thanks to our son Sohum, my husband Sumit, my father in-law Susanta for their patience and constant inspiration.

My sincere thanks are due to Dulal Mandal and Sayan Sengupta for helping me with MATLAB programs and output in Chapters 3 and 6. Mr Vijay Sharma, and my teacher and colleague Prof. Salil Kumar Sanyal, helped me in designing the hardware for QPSK modulation schemes. The valuable input and suggestions from Prof. Sanyal improved the quality of Chapter 5 for digital modulations.

I would also like to acknowledge my past and present graduate students for helping me in many ways. Special thanks to my students, Mr. Tamal Chakroborty and Mr. Arnab Raha who helped me in identifying the errors from the first edition and correcting the same. Thanks to Mr. Sibaram Khara, my former PhD student, for providing some of the figures for Chapter 15 related to integrated architecture. I would like to thank Mr. Pulak Roy, former MTech student for helping me in preparing the glossary of this book.

I would like to thank the following reviewers for their valuable comments and suggestions for improving the manuscript. I am indebted to so many references provided in this book for imparting knowledge for many topics which helped me immensely in writing this book.

Sanjeev Jain	*Rajasthan Technical University, Jaipur, Rajasthan*
Sudarshan Tiwari	*Motilal Nehru National Institute of Technology, Allahabad*
Brahmjit Singh	*National Institute of Technology, Kurukshetra*
Anita Seth	*Shri Govindram Seksaria Institute of Technology and Science, Indore*
Anjali Potnis	*Barkhatullah Institute of Technology, Bhopal*
Sumit Kundu	*National Institute of Technology, Durgapur*
Aniruddha Chandra	*National Institute of Technology, Durgapur*
T Rama Rao	*SRM University, Kancheepuram, Tamil Nadu*
N S V Shet	*National Institute of Technology, Surathkal*
Preeta Sharan	*Vemana Institute of Technology, Bangalore, Karnataka*

Students are always the source of inspiration to a teacher. I would like to thank all my students who have directly and indirectly helped, and influenced my writing of this book. It would be of great pleasure if this second edition helps the graduate and undergraduate students of our country in generating interest and knowledge thrust in wireless communications.

Last but not the least, sincere thanks to the editorial team at McGraw-Hill Education (India) for their support and cooperation in bringing out the present edition.

Iti Saha Misra

Feedback

Constructive criticism or suggestions are always welcome. Students can write to me at *itisahamisra@yahoo.co.in*

Publisher's Note

It will encourage us to receive your comments/compliments/ideas on this present edition. Please write to us at **tmh.ecefeedback@gmail.com** by mentioning the title and the author's name in the subject line. Report of any piracy related problems will be highly appreciated

VISUAL WALKTHROUGH

Introduction

Each chapter begins with an Introduction that gives a brief summary of the background and contents of the chapter.

Characteristics of Wireless Channels and Propagation Path Loss Models

4

Introduction

In wireless communications, the radio wave propagation medium largely affects the transmitted signal in between the path of transmitter and receiver. As the signal propagates through wireless channel, it experiences random fluctuations in time if the transmitter, receiver, or surrounding objects are moving, due to changing reflections and attenuation. Thus, the characteristics of the channel appear to change randomly with time. Unless, the channel response is known, it is difficult to design reliable systems with certain reliability and guaranteed quality.

The wireless propagation channel contains large obstacles scattering radio signal. A receiver may get many multipath signal components from various scattering objects those may interfere constructively or destructively depending on the phases of the multipath components and causing the phenomenon like fading. Further, the speed of the mobile device affects the signal level that also causes fading. Unlike free space propagation, scatterer introduces channel impairments like fading, multipath, delay spread, Doppler spread, etc. The signal fading may be frequency selective or non-frequency selective depending on the channel characteristics parameters. Thus, wireless channel is time variant in nature.

In this chapter, the wireless channel characteristics along with the radio propagation path loss models in wireless environments are discussed. The variation of received signal power over distances due to path loss and shadowing are characterized. Path loss is caused by dissipation of the power radiated by the transmitter as well as effects of propagation channel. Obstacles between the transmitter and receiver cause shadowing that attenuate received power through reflection, absorption, diffraction and scattering. Variation due to path loss occurs over large distance (few hundreds meters), whereas variation due to shadowing occurs over distances proportional to the length of the obstructing object (10–100 meters).

4.1 MULTIPATH PROPAGATION MECHANISMS

There are several mechanisms for creating multiple propagation paths. These are given below.

1. **Reflection:** Reflection occurs when a propagating electromagnetic waves falling on objects with smooth surface and large dimensions compared to wavelength.
2. **Diffraction:** Caused by obstructions with sharp irregularities in the path of the radio signal. Diffraction occurs when the radio path between the transmitter and receiver is obstructed by a dense body with large dimensions compared to λ, causing secondary waves to be formed behind the obstructing body. Diffraction is a phenomenon that accounts for RF energy traveling from transmitter to receiver without a line-of-sight path between the two. It is often termed shadowing because the diffracted field can reach the receiver even when shadowed by an impenetrable obstruction.
3. **Refraction:** The propagation path is diverted because of the difference in the electrical properties of the medium.
4. **Scattering:** Due to obstacles in signal path with much smaller dimensions than signal wavelength. These objects could be water droplets, clouds, or insects for example. Scattering causes the reflected energy to spread out (scatter) in all directions. In an urban environment, typical signal obstructions that yield scattering are lampposts, street signs, and foliage.
5. **Multipath:** Signals reflected through buildings, mountains, trees, open ground etc.

Sections and Subsections

Small-scale fading refers to the dramatic changes in signal amplitude and phase that can be experienced as a result of small changes (as small as a half-wavelength) in the spatial separation between a receiver and transmitter. Small-scale fading manifests itself in two mechanisms, namely, timespreading of the signal (or signal dispersion) and time-variant behavior of the channel. On a very short distance scale (comparable to one wavelength), if the received power fluctuates around a local mean value then the fading is said to be **small scale.** This is occurred due to the interference between the multiple components of the transmitted signal from different paths.

Small-scale fading is also called *Rayleigh fading* because if the multiple reflective paths are large in number and there is no line-of-sight signal component, the envelope of the received signal is statistically described by a Rayleigh probability distribution function. When there is a dominant non-fading signal component present, such as a line-of-sight propagation path, the small scale fading envelope is described by a Rician probability distribution function. A mobile radio roaming over a large area must process signals those experience both types of fading: small-scale fading superimposed on large-scale fading.

Table 4.1 *Different propagation mechanisms*

Cause	Effect
Multipath propagation	Fast fading, Delay Spread (Time dispersion)
Motion	Doppler Shift (frequency dispersion)
Shading	Slow fading
Signal Attenuation	Path Loss

4.3 CHARACTERIZATION OF THE CHANNEL

Channel characterization is an important topic in the investigation of mobile channel. Modeling the attenuation as a function of frequency, location and distance; time variation measurement of amplitude and phase; use of correlation for improved result, use of diversity techniques and to establish the channel impulse response to determine the delay spread are some of the usual way to address the channel characterization. Although channel fading is experienced as an unpredictable, stochastic phenomenon, powerful models have been developed that can accurately predict system performance. In this section, multipath channel is modeled due to the random time varying impulse response.

Linear Time Invariant (LTI) system does not have frequency component different from those of the input signal, thus no frequency shifts occur. But both nonlinear and time-varying systems introduce new frequency components other than those existing in the input signal. For wireless propagation environment, due to the mobility of mobile users and/or the surrounding scatterers, the channel is linear but time variant and is defined as Linear Time Variant (LTV) channel.

4.3.1 Time Varying Channel Impulse Response

Consider there are N number of scatterer or reflected objects in between the transmitter and receiver causing N multipaths received signals. We can express the received signal phasor as the sum of all possible multipath components within the receiver. Consider a narrowband signal $S(t)$, is transmitted at frequency f over the wireless channel,

$$S(t) = \text{Real of} \left\{ s(t)\, e^{j2\pi f t} \right\} \tag{4.1}$$

Where $s(t)$ is the complex envelop of the signal.

$$\text{The received signal } R(t) = \text{Real of} \left\{ \sum_{n=1}^{N} a_n(t) \times (t - \tau_n(t))^{\varphi 2\pi f (t - \tau_n(t))} \right. \tag{4.2}$$

where $a_n(t)$ is the amplitude variation of the different multipath received signal, $\tau_n(t)$ is the different time delay associated with multipath.

Each chapter neatly divided into sections and subsections under specific headings so that the subject matter is studied in a logical progression of ideas and concepts. Brief explanations will help readers grasp the topics.

Understanding wireless communication needs to understand wireless channel first. This book provides in depth coverage on that.

therefore independent as the channel is Gaussian random process. Thus the channel coherence time T_c is the time range over which $\Phi_H(\Delta t)$ is approximately non zero. Time varying channel de-correlates after time T_c. The function $S_H(\gamma)$ is called the Doppler power spectrum of the channel. The maximum γ value for which $|S_H(\gamma)|$ is zero is known as the Doppler frequency spread and is defined by f_d and it is the inverse of the coherence time T_c. Figure 4.27 is the representation of $S_H(\gamma)$ with respect to Doppler frequency change γ. The function $\Phi_H(\Delta t)$ is the autocorrelation function of a channel's response to a sinusoid and is known as the time-spaced function.

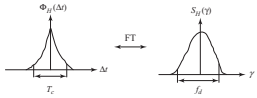


Fig. 4.27 *Doppler power spectrum, Doppler spread and coherence time*

The relationship between the scattering parameters and the four different channel response is given in Fig. 4.28.

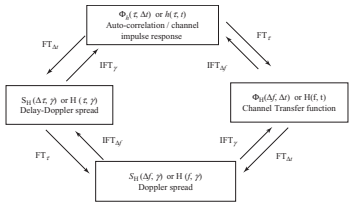


Fig. 4.28 *Relationship between the channel functions and Fourier transform*

Coherence Bandwidth: Defines a frequency range where the correlation coefficient is greater than a given threshold

$B_c \geq 1/2\pi\sigma_\tau$ where σ_τ is the rms delay spread

Coherence Time: The time duration over which two received signals has an amplitude correlation greater than a given threshold. It is the statistical measure of the time duration over which the channel impulse response is essentially invariant.

$T_c \approx 1/2\pi f_d$, where f_d is Doppler frequency shift

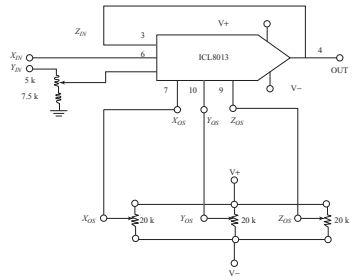


Fig. 5.16 *BPSK modulator circuit diagram using ICL8013*

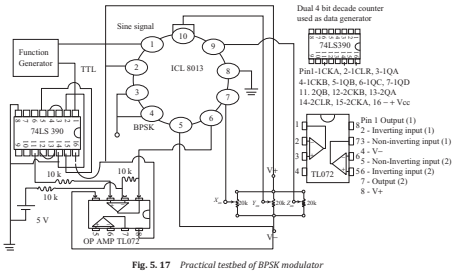


Fig. 5. 17 *Practical testbed of BPSK modulator*

The Op-Amp TL072 is used to convert the unipolar data bits to a bipolar signal as required by the analog multiplier ICL 8013 for proper multiplication to produce BPSK modulated signal. The principle of the unipolar to bipolar conversion circuit in our system is the level shifting of the data stream and amplification

Importance of various digital modulation techniques are clearly discussed with sufficient illustrations and mathematics for better understanding of the subject.

Evolutionary Path leading to technology development

The interface Mc works between the MGW and MSC servers. The interface Nc is either a classical SS7 interface or is IP based. For IP-based services, several other protocols need to be implemented.

The BICC (Bearer Independent Call Control) or SIP (Session Initiation Protocol) work between the MSC server and the GMSC server. The BICC supports narrowband ISDN service over broadband backbone network, ATM or IP.

Stream Control Transmission Protocol (SCTP) is necessary for transporting SS7 message on an IP interface. On the Mh interface, SS7 is implemented over SCTP/IP.

11.4 UMTS RELEASE 5

The primary aim of Release 5 is to provide IP-oriented services by the operators, unlike the core network modification of Release 4 for IP transport. It is expected to generate revenue by providing IP multimedia services. To meet the demand, conventional circuit-switched connections in the CS domain have to be replaced by enhanced IP-based access completed by packet switching. This is realized by the IP Multimedia Subsystem (IMS) that overlays the existing architecture. The IMS is an extention of packet-switched core network intended to offer access independence and inter-operation with wireline terminal access across the Internet. The interfaces therefore specified to conform as far as possible to IETF 'Internet Standards' where an IETF protocol is to be used. Fig. 11.5 shows the UMTS Release 5 network architecture using IMS.

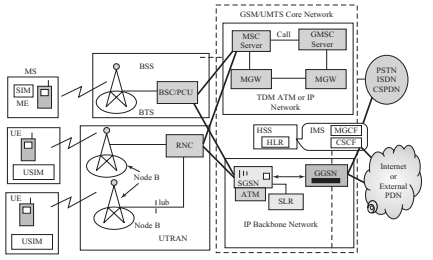


Fig. 11.5 *UMTS Release 5 networks*

The IMS uses PS domain to transport multimedia signaling and bearer traffic. For the future VoIP services, it is a prerequisite that all data of a multimedia service pass through the same core network. There are two components of IMS—one, Call State Control Function (CSCF) which is an entry point for signaling of incoming calls; and two, Media Gateway Control Function (MGCF), responsible for inter-working with the PSTN of CS domain. UMTS Release 5 introduces an integrated database called Home Subscriber Server (HSS), which provides the subscriber profile information. CSCF interacts with HSS for database queries for location management and routes the call accordingly.

The Session Initiation Protocol (SIP) has been standardised as the call and session control protocol for the IMS. SIP is a signaling protocol defined by IETF that establishes sessions, modifies and releases

Second-Generation Mobile Networks—GSM: Architecture and Protocols

8

Introduction

Global System for Mobile Communications (GSM) is a 2G digital mobile cellular system with its own communication protocols, interfaces and functional entities. Originally, GSM stood for 'Group Special Mobile' and intended to be a new telecommunication standard in Europe. The European Telecommunication Standards Institute (ETSI) was founded in 1988 and was responsible for standardizing the GSM technical specifications over Europe. The success of this standard has necessitated its renaming to Global System for Mobile Communications, which reflects its application worldwide. The first commercial GSM phone call was made in Finland on the 1st of July 1991.

The popularity and success of GSM can be summarized as follows:

- It is an open system standard, and anyone can have access to the specifications.
- It can give multi-vendor solutions because of a standardized interface.
- It can support roaming to users.
- It can maintain security and privacy of speech and data due to encrypted transmission.
- Digital transmission gives high speech quality and increases radio spectrum efficiency by the use of multiple access technology—FDMA, TDMA.
- Above all, it has an evolutionary implementation concept—upgradeable with downward compatibility.

ETSI has been standardized to operate on three basic frequency regions—900 MHz, 1800 MHz and 1900 MHz.

8.1 GSM NETWORK ARCHITECTURE

The GSM network architecture is given in Fig. 8.1. It consists of several base transceiver stations (BTS), which are clustered together and connected to a base station controller (BSC). Several BSCs are then

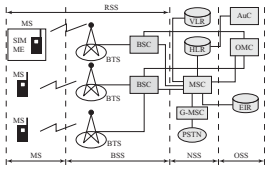
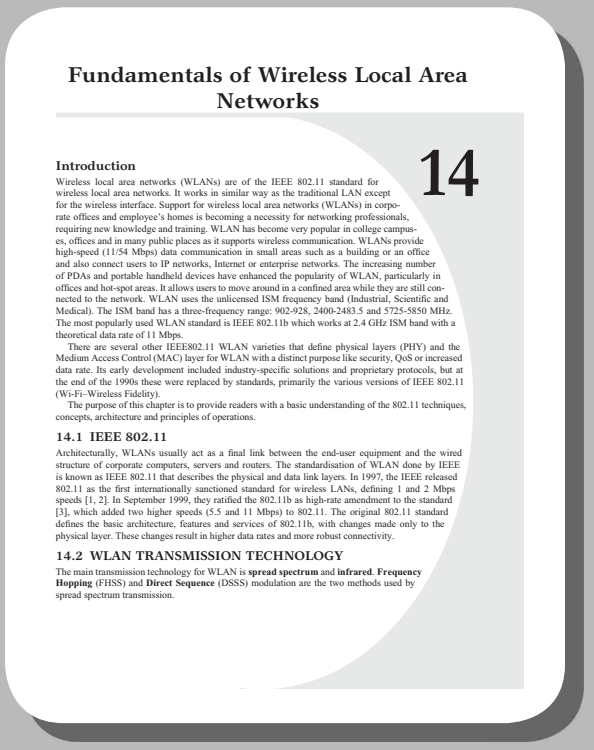


Fig. 8.1 *GSM network architecture*

Details of GSM protocol architecture, transition to GPRS packet data services and high data rate 3G UMTS networks will help the reader understand the technological revolution of mobile services. Follow through the path of evolution from cellular wireless to IP-based data networks through GPRS to UMTS a journey from 2G to 3G.

Chapter on IEEE 802.11 series

A comprehensive chapter is included to guide understanding, designing and operating of WLAN—the basic wireless platform in institutes and corporate houses.



Fundamentals of Wireless Local Area Networks

14

Introduction

Wireless local area networks (WLANs) are of the IEEE 802.11 standard for wireless local area networks. It works in similar way as the traditional LAN except for the wireless interface. Support for wireless local area networks (WLANs) in corporate offices and employee's homes is becoming a necessity for networking professionals, requiring new knowledge and training. WLAN has become very popular in college campuses, offices and in many public places as it supports wireless communication. WLANs provide high-speed (11/54 Mbps) data communication in small areas such as a building or an office and also connect users to IP networks, Internet or enterprise networks. The increasing number of PDAs and portable handheld devices have enhanced the popularity of WLAN, particularly in offices and hot-spot areas. It allows users to move around in a confined area while they are still connected to the network. WLAN uses the unlicensed ISM frequency band (Industrial, Scientific and Medical). The ISM band has a three-frequency range: 902-928, 2400-2483.5 and 5725-5850 MHz. The most popularly used WLAN standard is IEEE 802.11b which works at 2.4 GHz ISM band with a theoretical data rate of 11 Mbps.

There are several other IEEE802.11 WLAN varieties that define physical layers (PHY) and the Medium Access Control (MAC) layer for WLAN with a distinct purpose like security, QoS or increased data rate. Its early development included industry-specific solutions and proprietary protocols, but at the end of the 1990s these were replaced by standards, primarily the various versions of IEEE 802.11 (Wi-Fi–Wireless Fidelity).

The purpose of this chapter is to provide readers with a basic understanding of the 802.11 techniques, concepts, architecture and principles of operations.

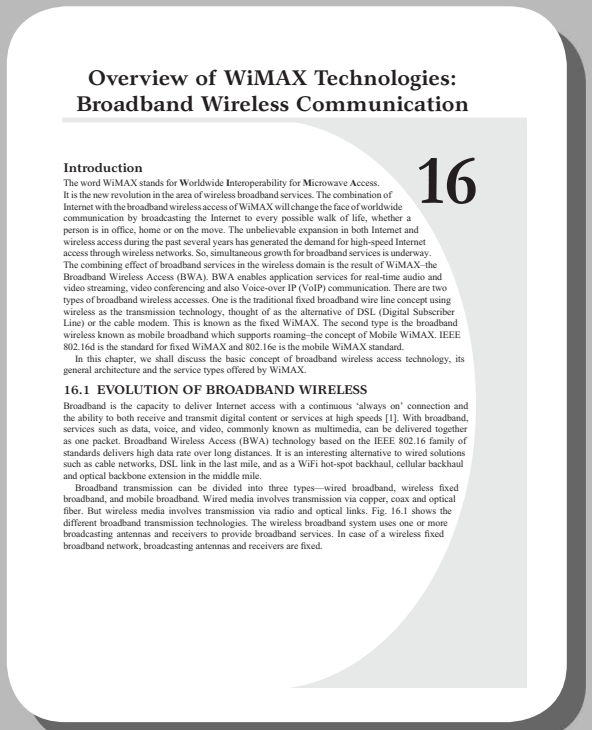
14.1 IEEE 802.11

Architecturally, WLANs usually act as a final link between the end-user equipment and the wired structure of corporate computers, servers and routers. The standardisation of WLAN done by IEEE is known as IEEE 802.11 that describes the physical and data link layers. In 1997, the IEEE released 802.11 as the first internationally sanctioned standard for wireless LANs, defining 1 and 2 Mbps speeds [1, 2]. In September 1999, they ratified the 802.11b as high-rate amendment to the standard [3], which added two higher speeds (5.5 and 11 Mbps) to 802.11. The original 802.11 standard defines the basic architecture, features and services of 802.11b, with changes made only to the physical layer. These changes result in higher data rates and more robust connectivity.

14.2 WLAN TRANSMISSION TECHNOLOGY

The main transmission technology for WLAN is **spread spectrum** and **infrared**. **Frequency Hopping** (FHSS) and **Direct Sequence** (DSSS) modulation are the two methods used by spread spectrum transmission.

Worldwide revolution of broadband mobile technology—WiMAX



Overview of WiMAX Technologies: Broadband Wireless Communication

16

Introduction

The word WiMAX stands for **W**orldwide **I**nteroperability for **M**icrowave **A**ccess. It is the new revolution in the area of wireless broadband services. The combination of Internet with the broadband wireless access of WiMAX will change the face of worldwide communication by broadcasting the Internet to every possible walk of life, whether a person is in office, home or on the move. The unbelievable expansion in both Internet and wireless access during the past several years has generated the demand for high-speed Internet access through wireless networks. So, simultaneous growth for broadband services is underway. The combining effect of broadband services in the wireless domain is the result of WiMAX–the Broadband Wireless Access (BWA). BWA enables application services for real-time audio and video streaming, video conferencing and also Voice-over IP (VoIP) communication. There are two types of broadband wireless accesses. One is the traditional fixed broadband wire line concept using wireless as the transmission technology, thought of as the alternative of DSL (Digital Subscriber Line) or the cable modem. This is known as the fixed WiMAX. The second type is the broadband wireless known as mobile broadband which supports roaming–the concept of Mobile WiMAX. IEEE 802.16d is the standard for fixed WiMAX and 802.16e is the mobile WiMAX standard.

In this chapter, we shall discuss the basic concept of broadband wireless access technology, its general architecture and the service types offered by WiMAX.

16.1 EVOLUTION OF BROADBAND WIRELESS

Broadband is the capacity to deliver Internet access with a continuous 'always on' connection and the ability to both receive and transmit digital content or services at high speeds [1]. With broadband, services such as data, voice, and video, commonly known as multimedia, can be delivered together as one packet. Broadband Wireless Access (BWA) technology based on the IEEE 802.16 family of standards delivers high data rate over long distances. It is an interesting alternative to wired solutions such as cable networks, DSL link in the last mile, and as a WiFi hot-spot backhaul, cellular backhaul and optical backbone extension in the middle mile.

Broadband transmission can be divided into three types—wired broadband, wireless fixed broadband, and mobile broadband. Wired media involves transmission via copper, coax and optical fiber. But wireless media involves transmission via radio and optical links. Fig. 16.1 shows the different broadband transmission technologies. The wireless broadband system uses one or more broadcasting antennas and receivers to provide broadband services. In case of a wireless fixed broadband network, broadcasting antennas and receivers are fixed.

WiMAX, the gateway of broadband services for mobile users, is covered precisely.

A step towards 4G networks—concept of integration

Cellular and WLAN Integration: Heterogeneous Network Architecture, Step towards 4G Networks

15

Introduction

Currently, various wireless technologies and networks exist that fulfill different needs and requirement of mobile users with respect to Quality of Service (QoS), radio coverage, multimedia service and data rate. Wireless LAN is a satisfactory solution for high-data-rate hot-spot access. Traditional and next generation cellular networks provide medium-data-rate wide-coverage service. Hence, integrating the complementary systems enables the best connection of mobile terminals anytime and anywhere. The discrepancy of access technologies and network architecture demand that the common infrastructure to integrate diverse wireless networks.

Nomadic users need to have ubiquitous access to remote information storage and computing services. As an evolutionary step toward 4G mobile communications, mobility in heterogeneous IP networks with GPRS, UMTS and IEEE 802.11 WLAN systems is seen as one of the central issues in the realisation of 4G communication network and systems.

Cellular (GPRS/3G) and Wireless Local Area Networks (WLAN) dominate wireless communication currently. Though, the cellular system is the main wireless media primarily for voice communication, it is recently being enhanced for data services. Contemporarily, the use of WLAN has increased in the hot-spot regions due to its high data rate and relatively low installation cost. So, integration of the two complementary systems is the focus for low-cost high-data-rate multimedia communications.

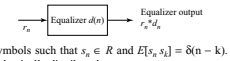
In 4G networks, to suffice the needs of users with relatively low mobility but yet to provide a substantial amount of bandwidth in hot spots, it becomes essential to integrate WLAN 802.11 systems with currently existing UMTS-GPRS systems. The integrated network can handle wider mobility of the mobile nodes (MN). The 3G cellular networks, e.g., UMTS are designed to provide users with voice and data services. The total cell capacity limits the per user data rate. Many a times high-speed requirements are clustered in small pockets. These clusters are termed **hot spots**. Network operators would like to employ efficient solutions, which can be easily integrated with their existing UMTS-based infrastructure. WLANs, therefore, offer an attractive solution. 3G cellular accesses based on code division multiple access (CDMA), either wideband CDMA or cdma2000, can be used to satisfy users who have a larger need for mobility, while 802.11 systems can be used to support users with much lesser coverage-area requirements. It is in light of this that the next wave of technological advance is already under consideration, i.e., 4G. Several international projects are already underway to ensure that services are suited to the characteristics of different delivery mechanisms, from cable networks to GPRS networks; and that the services can be delivered using the most appropriate of the range of networks available. Both the 3GPP and 3GPP2 experts are developing specifications for inter-working architectures that enable users to access their 2G and 3G data services from WLAN hot spots. In addition, the dynamic allocation of available spectrum between different wireless networks is also under investigation. The challenge is to design such

Covers the up-to-data convergence network concept of integration architecture and seamless roaming of users in heterogeneous networks.

Worked Examples

Example 6.4 Find the equalizer coefficients using 3-tap delay channel with response $h(n) = [h_{-1} = 0.2, h_0 = 0.9$ and $h_1 = 0.4]$ and $h(n) = [0\ 1.0\ .5]$ using minimum mean square error (MMSE) algorithm.

Solution Let the equalizer input in presence of AWGN is $r_n = y_n + w_n$ where w_n is the noise part.



Let s_n be the transmitted symbols such that $s_n \in R$ and $E[s_n\, s_k] = \delta(n - k)$. Also $E[s_n] = 0\ \forall\ n$. All the symbols are independent and identically distributed.

The signal received at the input of the equalizer is $r_n = y_n + w_n = \sum_{-1}^{1} h_k s_{n-k} + w_n$

Consider 3-tap equalizer with coefficients d_{-1}, d_0 and d_1.

The estimate of $s_n = \hat{s}_n = \sum_{-1}^{1} d_k r_{n-k}$

So the error is $e_k = s_n - \hat{s}_n$ and $\xi = E[e_k^2] = E[(s_n - \hat{s}_n)^2]$

The goal of MMSE equalizer is to minimize ξ.

Differentiating w.r.t d_q, $\partial\xi / \partial d_q = E[\partial / \partial d_q (s_n - \hat{s}_n)^2]$, where $q \in [-1, 0, 1]$

$$= E[\partial / \partial d_q (s_n - \sum_{-1}^{1} d_k r_{n-k})^2]$$

For MSE $\partial\xi / \partial d_q = 0$

$\partial\xi / \partial d_q = E[-2\, d_q r_{n-q} (s_n - \sum_{-1}^{1} d_k r_{n-k})] = 0$

as $d_q \neq 0$ for $\forall$ q, so

$E[r_{n-q} s_n - \sum_{-1}^{1} d_k r_{n-k} r_{n-q})] = 0$

$E[r_{n-q} s_n - \sum_{-1}^{1} d_k E(r_{n-k} r_{n-q})] = 0$

$\sum_{-1}^{1} d_k R_{rr}(q-k) = R_{sr}(q)$

$$\Rightarrow d_{-1} R_{rr}(q+1) + d_0 R_{rr}(q) + d_1 R_{rr}(q-1) = R_{sr}(q)$$

For $q = -1, 0, +1$ we get 3-equations

$d_{-1} R_{rr}(0) + d_0 R_{rr}(-1) + d_1 R_{rr}(-2) = R_{sr}(-1)$

$d_{-1} R_{rr}(1) + d_0 R_{rr}(0) + d_1 R_{rr}(-1) = R_{sr}(0)$

$d_{-1} R_{rr}(2) + d_0 R_{rr}(1) + d_1 R_{rr}(0) = R_{sr}(1)$

We know that the auto correlation is conjugate symmetric,

$R_{rr}(p) = R_{rr}(-p)$ for real p.

In matrix form,

$$\begin{bmatrix} R_{rr}(0) & R_{rr}(1) & R_{rr}(2) \\ R_{rr}(0) & R_{rr}(1) & R_{rr}(2) \\ R_{rr}(0) & R_{rr}(1) & R_{rr}(2) \end{bmatrix} \begin{bmatrix} d_{-1} \\ d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} R_{sr}(-1) \\ R_{sr}(0) \\ R_{sr}(1) \end{bmatrix}$$

Worked Examples are provided in sufficient number in each relevant chapter and at appropriate locations, to aid in understanding of the text material. Logical discussions are also given for better understanding.

Sufficient Illustrations

Illustrations are essential tools in books on engineering subjects. Ample illustrations are provided in each chapter to illustrate the concepts, functional relationships and in some cases illustrations are very self-explanatory

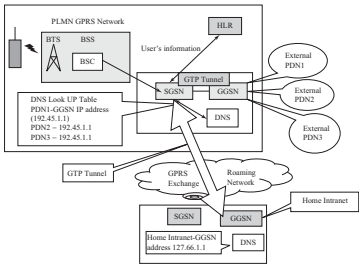


Fig. 9.4 *Illustration for PDP context*

The address assignment for PDP context may be static or dynamic. For a static case, the home PLMN (Public Land Mobile Network) of the MS permanently assigns the PDP address whereas in a dynamic case, the address assignment is done upon activation of PDP context by the operator's of the user, either in home PLMN or in visited PLMN. GGSN is responsible for the allocation and the activation/ deactivation of the PDP address allocation for a dynamic case. This address is used for packet forwarding between the Internet and the GGSN, and within the GGSN it is used to tunnel user data through the core network (SGSN to GGSN).

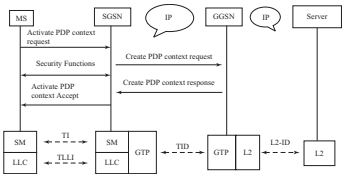


Figure. 9.5 *Steps for PDP context activation procedures*

inter-PLMN backbone are called Border Gateways (BGs). Among other things, they perform security functions to protect the private intra-PLMN backbones against unauthorized users and attacks.

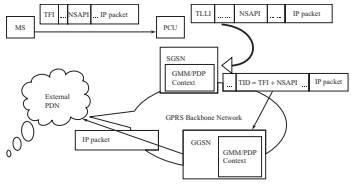


Fig. 9.6 *Transfer of data through GPRS network*

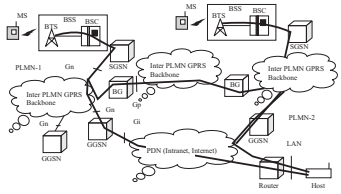


Fig. 9.7 *Example GPRS network for routing*

Let us now assume that the home-PLMN of the MS is PLMN2. The GGSN of PLMN2 has assigned an IP address to the mobile. Thus, the MS's IP address has the same network prefix as the IP address of the GGSN in PLMN2. The correspondent host is now sending IP packets to the MS. The packets are sent out onto the IP network and are routed to the GGSN of PLMN2 (the home-GGSN of the MS). The latter queries the HLR and obtains the information that the MS is currently located in PLMN1. It encapsulates the incoming IP packets and tunnels them through the inter-PLMN GPRS backbone to the appropriate SGSN in PLMN1. The SGSN decapsulates the packets and delivers them to the MS [2].

Online Resources and References

To effectively use the internet resources, references to relevant web addresses are provided in the text and a list of useful references are given at the end of each chapter.

LTE/EPC (called LTE-Advanced) that were determined in October 2010 to have successfully met all of the criteria established by the International Telecommunication Union Radio telecommunication Sector (ITU-R) for the first release of IMT-Advanced.

IMT-Advanced, which includes LTE-Advanced, provides a global platform on which to build next generations of interactive mobile services that will provide faster data access, enhanced roaming capabilities, unified messaging and broadband multimedia.

References

[1] *Cellular Phone History, Telephones, Mobile, Wireless,* www.affordablephones.net/HistoryCellula.htm
[2] *Waveguide–A Brief History of CellularTechnology*, www.wave-guide.org/archives/waveguide-3/cellular-history.html
[3] *Cellular Networks: Past, Present and Future,* www.acm.org/crossroads/xrds7-2/cellular.html
[4] MacDonald V.H., *The Cellular Concepts*, The Bell Systems Technical Journal, Vol. 58, No.1, pp. 15–43, Jan 1979.
[5] Mulder R.J., *DECT–A Universal Cordless Access System*, Philips Telecommunications Review, Vol. 49, No. 3, pp. 68–73, Sep 1991.
[6] Moralee D., *CT2–a New Generation of Cordless Phones*, IEE Review, pp. 177–180, May 1989.
[7] Oeting J., *Cellular Mobile Radio–An Emerging Technology*, IEEE Communications Magazine, pp. 10–15, Nov 1983.
[8] LeeW.C.Y., *Overview of Cellular CDMA*, IEEE Transaction on Vehicular Technology, Vol. 40, No. 2, May 1991.
[9] Correia L. and R. Prasad, *An Overview of Wireless Broadband Communications,* IEEE Communications Magazine, pp. 28–33, Jan 1997.
[10] Rappaport T.S., *Wireless Communications: Principles and Practice*, Pearson Education, 2002.
[11] 4G Mobile Broadband Evolution: 3GPP R10 and Beyond, February 2011 http://www.4gamericas.org

Questions

In each chapter, sufficient objective and descriptive questions are given. On an average there are 35 questions in each chapter totaling 35 × 13 = 455 questions. These are also very helpful to teachers in setting class work, assignments, quizzes and examinations.

Questions for Self-Test:

6.1 Describe the basic concept of equalization process to combat ISI effect in a fading channel.
6.2 What are the different types of equalization? Mention in which situation each of the equalizers would be useful.
6.3 Show that equalizer process enhances the noise.
6.4 Let the complex impulse response of the equalizer is $h_{eqz}(t) = \Sigma_n c_n \delta(t - nT)$, c_n is the complex filter coefficients of the equalizer. Show that in absence of noise, $F^*(-f)\, H_{eqz}(f) = 1$, where $F^*(-f)$ and $H_{eqz}(f)$ are the Fourier transform of the function $f^*(t)$ and $h_{eqz}(t)$ respectively.
6.5 Consider a linear transversal equalizer with $(2M+1)$ taps and tap coefficients w_{-M} to w_M, if $D(z)$ is the z-domain transfer function of the equalizer and $R(z)$ is the input of the equalizer coming from effective channel h_{eff}, then find the output of the equalizer $Y(z)$.
6.6 What is zero forcing equalizer? Explain the operation of ZF equalizer with effective channel response c_n and equalizer response d_n.
6.7 Consider the discrete frequency channel response in a multipath wireless propagation is $h(n) = [.1, .9, .3]$. Design the 3-tap zero forcing equalizer for the system for ISI free transmission. Plot the frequency domain channel and equalization response.
6.8 Repeat the Problem 6.6 for 5 tap and 7 tap ZF equalizers. Also plot the frequency response curves. Compare the results obtained in problems 6.6 and 6.7. Comment on the results.
6.9 Find the bit error probabilities for 3-, 5-, and 7- tap ZF equalizers. Consider the effective channel impulse response $h(n) = [1, .5] = \delta(n) + 0.5\,\delta(n-1)$.
6.10 Establish the relationship of Wiener-Hopf equation for MMSE equalizer.
6.11 Find the equalizer coefficients using 3-tap delay channel with response $h(n) = [h_{-1} = 0.1, h_0 = 0.9$ and $h_1 = 0.3]$ and $h(n) = [0\ 1.0\ .5]$ using minimum mean square error (MMSE) algorithm. Compare the results of the first part with respect to Problem 6. 6.
6.12 What is adaptive equalizer? Describe the basic principle of adaptive ness in respect of MMSE equaizer.
6.13 Plot the frequency response for MMSE equalizer using 3-, 5- and 7- tap channel with response $h(n) = [0\ 1.0\ .5]$ and $h(n) = [.1\ .9\ .3]$.

Appendices

Brief Overview of 3G LTE

Appendix – A

The successful evolution and deployment of GSM family of technologies, generally known as 3GPP family have gone through the development phase of GSM, EDGE, UMTS, HSPA, HAPA+, LTE and LTE-A. In the commercial market, HSPA+ continues its progress and in the way LTE revolution began.

Long Term Evolution (LTE) is the next step forward in cellular 3G services. In the world of telecommunications, people today are more mobile and connecting themselves to mobile interfaces than ever. We have more new sophisticated handheld mobile devices to stay in touch with one another over wireless networks. The high speed mobility and on demand access of network and multimedia applications is the driving force for LTE technology development. LTE is an important technology transfer starting from circuit switch to packet switch and landed over to All-IP network architecture.

At the beginning of mobile broadband, AT&T launched UMTS enhanced with High Speed Downlink Packet Access (HSDPA) in 16 major markets throughout the U.S. in December 2005, and has become the first operator in the world to launch HSDPA on a wide-scale basis. AT&T deployed HSPA – capable of peak theoretical downloads speeds of up to 3.6 Mbps – in more than 350 U.S. cities and then upgraded its entire HSPA network to peak theoretical capabilities of up to 7.2 Mbps. AT&T announced plans to deploy HSPA+ in 2010, and began trials of LTE in the 700 MHz band with commercial deployment of LTE in and around 2011. The useful application of LTE is Voice Over LTE (VoLTE) expected to be available in the year 2013.

LTE supports both frequency-division duplex (FDD) and time-division duplex (TDD). In order to support large number of different spectrum allocations, LTE supports also a wide range of system bandwidths. It also aims for a smooth evolution from earlier 3GPP systems such as time division synchronous code division multiple access (TDSCDMA) and wide-band code division multiple access/high-speed packet access (WCDMA/HSPA), as well as 3GPP2 systems such as code division multiple access (cdma) 2000 [A.1].

The LTE as defined by the 3rd Generation Partnership Project (3GPP) is a highly flexible radio interface, which is using the VoIP to transmit the voice services and packet the data for all services. The first release of LTE was published in March 2009 and is referred to as LTE Release-8. LTE Release- 8 provides high peak data rates of 300 Mb/s on the downlink and 75 Mb/s on the uplink for a 20 MHz bandwidth. In LTE Release 8, orthogonal frequency-division multiplexing (OFDM) is the DL (Downlink) multiple access scheme, while single-carrier frequency-division multiple access (SC-FDMA) is the UL (uplink) multiple access scheme. LTE Release-8 also supports scalable bandwidth up to 20 MHz, and uses DL/UL frequency selective and DL frequency diverse scheduling, respectively. The DL subframe structure is common to both TDD and FDD. The Media Access Control Layer (MAC layer) in data link layer of Open System Interconnection (OSI) aims to control the authority of user about the accessing media and resource.

Overview of Bluetooth Technology

Appendix-B

Bluetooth is the short-range radio link technology developed with the intention to replace the cable connecting portable and/or fixed electronic devices. Bluetooth allows for the replacement of the many propriety cables that connect one device to another with one **universal radio link**. Bluetooth radio modules operate in the unlicensed ISM band at 2.4GHz, and avoid interference from other signals by hopping to a new frequency after transmitting or receiving a packet. The beauty of this technology is robustness, low complexity, low power and low cost and designed to operate in noisy frequency environments.

The research on Bluetooth was initiated at Ericsson of Sweden in 1994. The idea of Bluetooth comes from the desire to connect cellular phones with other devices without a cable. The name "Bluetooth" came after the 10th century Viking king of Denmark Herald Bluetooth.

Bluetooth technology is a short-range communications technology that is simple, secure, and everywhere ranging from mobile phones and computers to medical devices and home entertainment products. It is also low cost and consumes low power. Two Bluetooth enabled devices connecting to each other formed a pair. Any Bluetooth enabled device, almost everywhere in the world, can connect to other Bluetooth enabled devices located in proximity to one another.

Connections between Bluetooth enabled electronic devices allow these devices to communicate wirelessly through short-range, ad hoc networks known as piconets. A Piconet session is a communication link that must be created between devices to communicate with each other. Piconets are established dynamically and automatically as Bluetooth enabled devices enter and leave radio proximity. Each device in a piconet can also simultaneously communicate with up to seven other devices within that single piconet and each device can also belong to several piconets simultaneously.

A fundamental strength of Bluetooth wireless technology is the ability to simultaneously handle data and voice transmissions, which provides users with a variety of innovative solutions such as hands-free headsets for voice calls, printing and fax capabilities, and synchronization for PCs and mobile phones etc.

The range of Bluetooth technology is application specific. The Core Specification requires a minimum range of 10 meters/ 30 feet, but there is no set limit and manufacturers can tune their implementations to provide the range needed to support the use cases for their solutions.

If two devices come in contact with each other within 30 feet, the user will be prompted to initiate a communication session. Users then can either deny or accept the request to initiate a session. Only devices approved by the user can take part in the session. Data will appear as noise to unauthorized devices providing a great security feature.

A very brief understanding of digital communication in the context of wireless communications is provided that will help readers to get through the overview of digital technology revolution used in different wireless systems.

Some sample questions are provided to help the teachers.

Introduction to Wireless Communications and Networks

1

Wireless and mobile communications provide the ability to communicate with people on the move. Wireless communication systems and networks evolved through generations, from the first to the fourth. When commercial mobile telephony began in the 1940s, digital wireless and cellular communications slowly entered into the picture. Use of vacuum tubes and transistors made the evolution of early telephone networks possible. Similarly, the wireless revolution started after low-cost microprocessors and digital switching became available. Further, with the advancement of digital technologies and large-scale integration for miniaturization of radio frequency circuits, portable radio handheld devices become cheaper and readily available. These helped users to have easy access to wireless communications.

Mobile radios began operating at 2 MHz above the present AM radio broadcast band in 1921 in the United States [1]. The Detroit Michigan Police Dept. made the earliest significant use of a mobile radio in a vehicle in the United States [2]. These were mainly experimental police-department radios, with practical systems not implemented until the 1940s. Primarily, a mobile radio was used for police and emergency services with little thought given to private mobile telephone use. The channels soon became overcrowded.

In the 1940s, new frequencies between 30 and 40 MHz were made available. Shortly thereafter, many individuals, companies, and public agencies purchased and operated their own mobile units. In the year 1945, AT&T and Southwestern Bell introduced the first commercial mobile telephone system, launched in St Louise, Missouri, US, with three channels at 150 MHz and 60 KHz channel spacing allocated by FCC (Federal Communications Commission). But there was a problem with the mobile equipments that were interference limited. A public mobile system using frequencies in the 35 to 44 MHz band began operating along the highway between New York and Boston. These early mobile telephone systems used the push-to-talk operation [2]. The Improved Mobile Telephone System (IMTS) operating at 450 MHz band with direct dialing, automatic channel selection capability became the standard for mobile telephone service in US in the year 1969, and was introduced by Bell Systems. It was a full-duplex mode of operation in a two-way communication path like telephone systems.

Cellular wireless mobile services are a high capacity system of providing direct dialing between automobiles and other forms of portable telephones. It was first made available in the United States at 1984, and was followed by tremendous growth within a few years. The basic principle of such services is to operate at small power and bypass copper wires–hence, they are called wireless loops. Cellular communication, also called PCS (Personal Communication Services) aims to reach a person anywhere and anytime to provide telephone services.

Prior to cellular mobile telephone service, mobile telephony was conducted by two-way radio systems that allow only a few number of users over a service area. A single high-power transmitter located centrally was used over a 50-mile diameter. The cost was too high, and service was provided by an independent company. With more users coming in, these systems failed. Thus, the concept of cellular telephony came into being. The basic principles of mobile cellular services were introduced by Bell Labs in 1940, but were not publicly available and commercially viable until the 1970s [1]. Cellular mobile telephony took a long time to be used for commercial utilization because of the federal

regulatory process. In the 1960s, the new IMTS, launched by Bell Systems, brought about many improvements like direct dialing and higher bandwidth. The first analog cellular systems were based on IMTS and were developed in the late 1960s and early 1970s. The systems were cellular because the geographical coverage areas were split into smaller areas known as 'cells', each of which was served by a low-power transmitter and receiver.

AT&T had applied to FCC for granting permission to introduce the Advanced Mobile Phone Systems (AMPS) based on the cellular principle. In 1975, FCC reallocated a portion of the UHF television band for the use of cellular communication. In March 1977, FCC granted authorization to the Illinois Bell Telephone Company to install and test the first development of AMPS in Chicago [4]. It was a successful attempt and in 1983, it offered the first cellular service in United States. The delay caused was due to the FCC which took a long time to simulate the regulatory action.

Cordless telephones, prior to cellular telephony, worked within 100 feet, enabling a user to walk about the home with the telephone. The cordless set was a two-way radio and was connected to the standard telephone line. With a wireless handset, a cordless telephone communicates with a base station connected to a fixed telephone landline via radio waves and can only be operated close to its base station within a range of 100 metres. Unlike a standard telephone, a cordless telephone needs household mains electricity to power the base station. The cordless handset is powered by a battery. It is recharged by the base station when the handset is connected it and not in use. Initially, it was analog, but modern cordless telephone systems use a digital sophisticated technology. The main problem of a cordless system is that it does not provide hand-off facility. The second-generation cordless telephone was designed in the United Kingdom in 1989 and was based on digital technology that supports circuit-switched voice services. Digital European Cordless telecommunications (DECT) was designed by the ETSI (European Telecommunications Standards Institute) in the year 1992 [5].

The first generation (1G) mobile cellular system started in the 1980s. It was based on analog radio transmission and circuit-switched techniques. At that time, there was no worldwide or even Europe-wide coordination for the development of technical standards for the system. The main mobile services provided by 1G were circuit-switched voice communications. 1G used analog FM for speech transmission. The individual calls used different frequencies and share the available spectrum through FDMA. 1G lacked the ability to support roaming between different network operators and countries. 2G, which is completely digital, employing either TDMA or CDMA, began to emerge in the early 1990s. The advantages of digital cellular technologies over analog cellular networks include increased capacity and security. Technology options such as TDMA and CDMA offer more channels in the same analog cellular bandwidth and encrypted voice and data.

2G brought a number of significant advancements over 1G wireless networks. Standards for core networks were introduced to support roaming between network operators and countries. In addition to circuit-switched voice services, 2G enabled the first wave of mobile data and mobile Internet services, now widely adopted by users.

In the late 1990s standardization efforts for third-generation (3G) wireless networks began to increase radio system capacities and per user data rates over 2G systems. 3G activities were limited in Europe and North America under the respective names of IMT-2000 (DS-CDMA—Direct Sequence Code Division Multiple Access) and CDMA-2000 (MC-CDMA—Multi Carrier CDMA). Its aim was to support IP-based data, voice and multimedia services and a broader range of IP-based mobile services.

In general, the word 'mobile' is defined as a fast-moving handset (radio terminal) having wireless network access, such as the cellular telephone system. In a cellular mobile communication system, a radio cell is defined as the geographical area served by a base station through which mobile users can communicate. Variable power levels allow cells to be sized according to the subscriber density and demand within a particular region.

To increase capacity, cellular wireless systems use frequency reuse techniques. Channels (frequencies) used in one cell can be reused in another cell some distance away. Cells can be added to accommodate growth, creating new cells in unserved areas or overlaying cells in existing areas. During roaming, mobile users change the base station from cell to cell. This generates the requirement of hand-off for uninterrupted services. The mobility of the users necessitates location management and hand-off management, together known as mobility management, of the wireless network systems. To support a higher transmission rate, the scarce radio spectrum needs to be reused more and more, thus reducing the cell size. This again generates more probable hand-off for a mobile user. So, a trade-off is required to reduce the management overhead of the wireless networks for mobility management.

In a cellular system, a base station can communicate with mobiles via a radio channel as long as they are within range. The channel is made of two frequencies, one for transmitting to the base station (uplink) and one for receiving information from the base station (downlink). Radio energy dissipates over distance; so the mobiles must be within the operating range of the base station. The radio channel suffers many propagation channel impairments during transmission, like multipath delay spread, fading, and intracell and intercell interferences. Proper knowledge of these phenomena is needed for good transmission and reception reducing interference and noise.

Depending on the coverage area, wireless communication may be categorized as wireless local area networks within indoor coverage and public wide area communication. The first type is mainly standardized by IEEE 802.11 wireless LAN and a promising part of the high data rate wireless access networks. IEEE 802.11 consists of a family of standards that defines the physical layers (PHY) and medium access control layer (MAC), architecture and protocols, security, QoS and mobility issues of wireless local area networks. The second category is cellular wireless communication that has progressed through 1G, 2G, 2.5G to 3G to increase the radio system capacities and per user data rates over 2G, to support IP-based data, voice and multimedia services, and at the same time improve interoperability and QoS.

By seamlessly integrating the services over a wide area cellular wireless networks with the WLANs, mobile operators and service providers may get full advantages of both the networks. But integration between heterogeneous networks brings many challenges, like vertical handoff, for efficient mobility management.

The future trend for mobile wireless communication is to extend the networks to access the Internet for mobile data transport. Use of IP for mobile users will become the basis for next generation data networks. Thus, the Internet, in the use of broadband wireless systems like WiMAX, will predominate the next generation of web services worldwide.

The fourth generation (4G) networks are becoming the most discussed topic nowadays. The fast technology advancement has generated a hope worldwide, both among the wireless operators and customers, to fulfill the demand of low cost, high-speed video and other forms of high-speed data over the 4G network platform. Evolution of 4G will be driven by better service quality, heterogeneity of networks, fast response, and high session rate. The broadband wireless services may be dominated by the use of WiMAX and its variants. The complementary services of both cellular 2G/3G with the WiFi and WiMAX within a unified IP-based platform may realize the 4G dreams within a short period.

Long Term Evolution (LTE) is the next step forward in cellular 3G services. LTE has been identified as a new wireless standard by the 3rd Generation Partnership Project (3GPP) which is using the VoIP to transmit the voice services and packet the data for all services. LTE provides an uplink speed of up to 50 megabits per second (Mbps) and a downlink speed of up to 100 Mbps. It provides higher bandwidth, lower latency, and better QoS. LTE also capable of proving scalable bandwidth1.25 MHz to 20 MHz that will bring many technical benefits to cellular networks.

4G Mobile Broadband Evolution *3GPP Release 10 and Beyond - HSPA+, SAE/LTE and LTE-Advanced*, provides detailed discussions of Rel-10 including the significant new technology enhancements to

LTE/EPC (called LTE-Advanced) that were determined in October 2010 to have successfully met all of the criteria established by the International Telecommunication Union Radio telecommunication Sector (ITU-R) for the first release of IMT-Advanced.

IMT-Advanced, which includes LTE-Advanced, provides a global platform on which to build next generations of interactive mobile services that will provide faster data access, enhanced roaming capabilities, unified messaging and broadband multimedia.

References

[1] *Cellular Phone History, Telephones, Mobile, Wireless,* www.affordablephones.net/HistoryCellula.htm
[2] *Waveguide–A Brief History of CellularTechnology*, www.wave-guide.org/archives/waveguide-3/cellular-history.html
[3] *Cellular Networks: Past, Present and Future,* www.acm.org/crossroads/xrds7-2/cellular.html
[4] MacDonald V.H., *The Cellular Concepts*, The Bell Systems Technical Journal, Vol. 58, No.1, pp. 15–43, Jan 1979.
[5] Mulder R.J., *DECT–A Universal Cordless Access System*, Philips Telecommunications Review, Vol. 49, No. 3, pp. 68–73, Sep 1991.
[6] Moralee D., *CT2–a New Generation of Cordless Phones*, IEE Review, pp. 177–180, May 1989.
[7] Oeting J., *Cellular Mobile Radio–An Emerging Technology*, IEEE Communications Magazine, pp. 10–15, Nov 1983.
[8] LeeW.C.Y., *Overview of Cellular CDMA*, IEEE Transaction on Vehicular Technology, Vol. 40, No. 2, May 1991.
[9] Correia L. and R. Prasad, *An Overview of Wireless Broadband Communications,* IEEE Communications Magazine, pp. 28–33, Jan 1997.
[10] Rappaport T.S., *Wireless Communications: Principles and Practice*, Pearson Education, 2002.
[11] 4G Mobile Broadband Evolution: 3GPP R10 and Beyond, February 2011 http://www.4gamericas.org

Evolution of Modern Mobile Wireless Communication Systems

2

Introduction

Over the recent past, we have witnessed a considerable amount of growth in the wireless industry, both in terms of mobile technology and subscribers. Both network operators and vendors have realised the importance of efficient networks with equally efficient design processes. Wireless communications have become pervasive. The increasing number of mobile phones, personal digital assistants (PDA), mobile handheld devices and mobile subscribers have necessitated the upgradation of cellular technology through several generations to support demand for modern data and multimedia services along with voice communications.

With all the technological advances, and the simultaneous coexistence of 2G, 2.5G and 3G networks, many new design scenarios have developed, and the inter-operability of the networks has to be considered. Mobile networks are differentiated from each other by the word 'generation' such as first, second, etc. There is a big generation gap between the successive technologies.

The first-generation analog wireless networks were meant for voice communications. As the need grew to transmit data and multimedia services through wireless networks, the evolution of public mobile services—evolving rapidly from text-based instant messaging to mobile Internet services like email, commercial transactions (banking, reservation, credit and billing enquiry)—, happened due to the evolution of wireless communication systems, where the features provided for broadband digital communication were incorporated.

Wireless networks include local, metropolitan, wide, and global areas in order to fit different coverage areas and communication needs. Depending on the coverage, wireless networks are categorized into the following:

1. Personal Area Networks (PAN)
2. Wireless Local Area Networks (WLAN)
3. Low-tier wireless systems
4. Public wide area cellular systems

From an engineering perspective, development of wireless systems has been struggling with different fundamental design bottlenecks, each of them typical to a particular phase of development. The solution of one problem leads to a technological jump, but also invites another new problem.

In this chapter, we will cover the evolution path of such networks in the context of service needs along with their basic principles of operations.

2.1 PERSONAL AREA NETWORK: PAN

PAN is basically a low-power, short-range radio trans-receiving system that allows a user or device to communicate with another user or device within the coverage area of 10 m to 50 m within a room or vehicle. **Bluetooth** is the example of such communication. It is a short-range radio wave wireless technology operating in the 2.4 GHz frequency spectrum. With an operating range of 30 feet (10 metres) and a maximum transmission rate of a mere 1 Mbps, the Bluetooth wireless technology was created to replace the cables used on mobile devices with radio frequency waves. The technology encompasses a simple low-cost, low-power, global radio system for integration into mobile devices. Such devices can form a quick ad-hoc secure pico-net and communicate among the

connected devices. This technology creates many useful mobile usage models because the connections can occur while mobile devices are being carried in pockets and briefcases.

Even the IEEE organisation has recognised the need for wireless cable replacement technology and started the development of the 802.15-working group that focuses on this market for Wireless Personal Area Networks. This specification is based on the Bluetooth technology. IEEE 802.15 is also defining short-range radio communication under PAN with a high data rate of up to 20 Mbps. Laptop computers and PDAs may form PAN within a meeting room connecting the devices like printers and projectors wirelessly in an ad-hoc manner. Fig. 2.1 shows an example of a PAN.

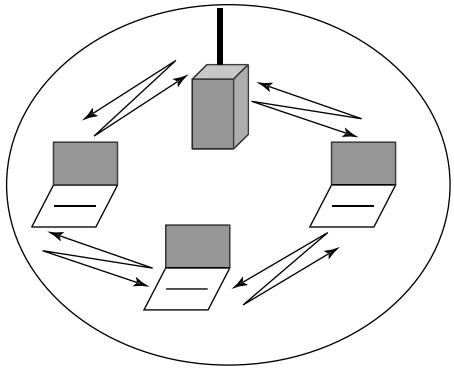


Fig. 2.1 *Personal area network*

IEEE 802.15 IEEE 802.15 is a short-range wireless communication standard that supports WPAN. It uses the low-power, low-cost and small-sized networks for which communication is within the range of user surroundings and between handheld devices, PDAs and computer peripherals. IEEE 802.15.3 [8] standard was developed for high-speed WPAN with a 55 Mbps data throughput within a coverage range of 10 m. These networks operate in a 2.4 GHz unlicensed band.

2.2 LOW-TIER WIRELESS SYSTEM

The coverage area of low-tier systems is less than 500 m in the outdoors and less than 30 m in the indoors. This system connects a telephone handset wirelessly with the base station that is connected to the fixed-line wired telephone network. Users with low mobility (pedestrians) are supported through this system. Some of the standards of this category are second-generation Cordless Telephones (CT2), Digital European Cordless Telecommunications (DECT), Personal Access Communication Systems (PACS) and Personal Handyphone Systems (PHS).

2.2.1 Cordless Telephone—Second Generation (CT2)

A cordless telephone, or portable telephone, is a telephone with a wireless handset which communicates with a base station connected to a fixed telephone landline via radio waves and can only be operated close to its base station, typically within less than 100 metres, such as in and around the house. Unlike a standard telephone, a cordless telephone needs the household mains electricity to power the base station. The cordless handset is powered by a battery which is recharged by the base station when the handset is connected to the base station and is not in use. A cordless telephone set is shown in Fig. 2.2.

The second-generation cordless telephone was designed in United Kingdom in 1989 and is based on digital technology that supports circuit-switched voice services. In the United States, there are seven frequency bands that have been allocated by the Federal Communications Commission (FCC) for uses that include cordless phones. These are:

1. 1.7 MHz
2. 27 MHz (allocated in 1980)
3. 43–50 MHz (allocated in 1986)
4. 900 MHz (902–928 MHz) (allocated in 1990)

Fig. 2.2 *Modern cordless telephone system*

5. 1.9 GHz (1920 N–1930 MHz) (allocated in October 2005)
6. 2.4 GHz (allocated in 1998)
7. 5.8 GHz (allocated in 2003)

All telephones in the US use the 900 MHz, 2.4 GHz, or 5.8 GHz bands.

2.2.2 Digital European Cordless Telecommunications (DECT)

DECT was designed by the ETSI (European Telecommunications Standards Institute) in the year 1993. This is primarily used for domestic or corporate purposes. DECT can also be used for wireless data transfers. It runs by Private Branch Exchanges (PBXs). It supports both circuit-switched voice and data services. DECT is a GSM-like cellular system. A major difference between the systems is the cell radius—DECT cells have a radius of 25 to 100 metres, while GSM cells are of 2 to 10 km radii.

Some of the DECT operating features are:

1. Bit rate—32 kbit/s
2. Frequency—1900 MHz
3. Carriers—10 (1880–1900 MHz)
4. Timeslots—2×12 (up and downstream)
5. Channel allocation—dynamic
6. Traffic density—10000 Erlangs/km^2

The DECT physical layer uses Frequency Division Multiple Access (FDMA), Time Division Multiple Access (TDMA) and Time Division Duplex (TDD). This means that the radio spectrum is divided into physical channels in two dimensions, both in frequency and time.

The DECT media access control layer controls the physical layer and provides connection-oriented, connectionless and broadcast services to the higher layers. It also provides encryption services with the DECT Standard Cipher.

There are four application areas of DECT:

1. Domestic DECTs are connected to a base station that connects the PSTN (public switched telephone network). More than one DECT handset can be connected to the base station.
2. Business DECTs connected to PBX with many radio-fixed parts. The DECT handsets dynamically connect to these base stations and support handover feature.
3. Public DECT connected to the PSTN, though rare, can be an alternative to GSM at high deployment density.
4. Local loop (though, very rare), in this case, a DECT radio link replaces the normally wired connection between the final PSTN distribution point to the subscriber.

DECT was developed by ETSI but has since been adopted by many countries all over the world. DECT is used in all countries in Europe. Outside Europe it is used in most parts of Asia, Australia and South America.

2.2.3 Personal Handyphone System (PHS)

PHS is also marketed as the **Personal Access System** (**PAS**). It is a cordless telephone which works in a mobile environment operating in the 1880–1930 MHz frequency band. PHS also supports handover from one cell to another. The PHS cells are small, having area of coverage ranging from 10 m to 100 m. PHS is suitable to deploy in dense urban areas, in contrast to rural areas that are mainly supported by GSM (Global System for Mobile) covering multi-kilometre ranges. The small cell size also makes it difficult to make calls from rapidly moving vehicles due to handoff.

PHS was originally developed by NTT in Japan in the year 1989 to compete with GSMs like the PDC (Personal Digital Cellular) system, but failed to gain importance due to its limited coverage and roaming facility.

PHS uses TDMA/TDD as its radio interface and was designed to support a channel rate of 384 kbps which is higher than the data rates of other low-tier systems. Modern PHS phones can support value-added services such as high-speed Internet connection, emailing and text messaging.

2.2.4 Personal Access Communications Systems (PACS)

PACS was designed in the United States by Telcordia, and then by Bellcore in the year 1992 to provide wireless access to local exchange carriers. The radio coverage of this system is about 500 m. It supports voice, data and even video for use in indoor and outdoor microcells.

2.3 PUBLIC WIDE-AREA WIRELESS NETWORKS

Public commercial wide-area mobile networks provide services over large geographical areas supporting both pedestrian and vehicular speed users. Wide-area radio systems have made their way through several generations with the technology upgradations from one generation to another with varied network capabilities—the first-generation (1G) analog, voice-only communications, to the second-generation (2G) digital, voice and data communications, and further to the third-generation (3G) wireless networks as a convergence of wireless and the Internet.

2.3.1 First-Generation (1G) Wireless Networks

The first-generation (1G) mobile cellular system started in the 1980s. It was based on analog radio transmission and circuit-switched techniques. At that time, there was no worldwide or even Europe-wide coordination for the development of technical standards for the system. The main mobile services provided by 1G were circuit-switched voice communications. 1G used analog FM for speech transmission. The individual calls used different frequencies and shared the available spectrum through FDMA.

The following are the types of 1G radio system standards in the world:

1. The Nordic countries deployed the Nordic Mobile Telephones or NMTs
2. The United Kingdom and Ireland went for the Total Access Communication System or TACS
3. Advanced Mobile Phone systems (AMPs) in North America
4. NTT-Nippon Telephone and Telegraph in Japan

AMPS was first deployed in 1983 in Chicago, USA, using the 800-MHz to 900-MHz frequency band and the 30-kHz bandwidth with 832 channels for each channel with a transmission rate of 10 kbps. Transmissions from the base stations to mobiles occur over the forward channel using frequencies between 869–894 MHz. The reverse channel is used for transmissions from mobiles to the base station, using frequencies between 824–849 MHz. The smallest reuse factor that would fulfill the 18-dB signal-to-interference ratio (SIR) using 120-degree directional antennas was found to be 7. Hence, a 7-cell reuse pattern was adopted for AMPS. It is the first standardised cellular service in the world and is currently the most widely used standard for cellular communications, such as the United States, South America, China, and Australia. The key features of AMPS can be summarized as the following:

1. 25-MHz band in each uplink, from 824 to 849 MHz
2. 25-MHz band in each downlink, from 869 to 894 MHz
3. AMPS uses a channel spacing of 30 kHz with a total capacity of 832 channels
4. Supports a data-transmission rate of 10 kbps

TACS was introduced in Europe on the 900 MHz frequency band, allowing up to 1320 channels using a 25-kHz channel spacing. Both AMPS and TACS use the frequency modulation technique for radio transmission, but frequency division multiplexing (FDMA) is used for providing access to the user. The TACS are now obsolete in Europe, having been replaced by the second generation, more scalable and all-digital Global System for Mobile Communication (GSM) system. The key features of GSM can be summarized as

1. 25-MHz band in uplink from 890 MHz to 915 MHz
2. 25-MHz band in down-link from 935 MHz to 960 MHz
3. Channel spacing of 25 kHz
4. Total capacity of channels—1000
5. Supports a data-transmission rate of 8 kbps

The Nordic Mobile Telephony (NMT) is a classic cellular standard using a 13.5-kHz channel spacing developed by Ericsson, and being used in 30 countries around the world. It is the common Nordic standard for analog mobile telephony as established by the telecommunications administrations in Sweden, Norway, Finland and Denmark in the early 1980s.

Finally, NTT, Nippon Telephone and Telegraph system was deployed in Japan with the following features:

1. 15-MHz band in the uplink from 925 to 940 MHz
2. 15-MHz downlink from 870 to 885 MHz
3. Channel spacing of 25 kHz, 600 channel of capacity
4. Data-transmission rate of 0.3 kbps

It is to be observed that all the 1G systems differ by features such as the location of the spectrum band, channel capacity, spacing and also with data rate. The basic 1G network architecture is given in Fig. 2.3.

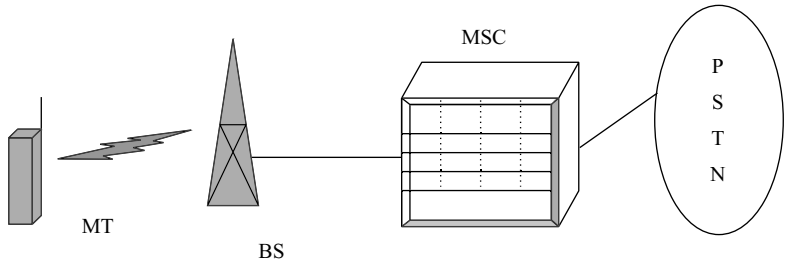


Fig. 2.3 *First-generation networks architecture*

In general, a wide-area wireless network typically consists of a Radio Access Network (RAN) to which the mobile terminal (MT) is wirelessly connected in order to access the core network (CN). RAN consists of wireless Base Stations (BS), and the *BS provides radio coverage over a geographical area called a cell*. These radio cells are typically arranged in an array using frequency-reusing technique for increased spectrum efficiency. So, wide-area wireless systems are usually known as cellular systems.

The other part of the system is the core network (CN). CN is a wire-line network used to interconnect RAN, and RAN to other networks, such as PSTN. In this way, MT gets wider coverage. Mobile Switching Center (MSC) is the main entity for CN.

The main problem of 1G systems was roaming between different network operators, as each operator operated in a proprietary core network. Further, 1G standards were used in different countries; it was impossible for a user to roam from one country to another. Also, efficient use of the frequency spectrum was not possible.

2.3.2 Second-Generation (2G) Wireless Cellular Networks

In the early 1990s, the second generation (2G) cellular systems began to evolve with a number of significant advancements over 1G wireless systems. Second generation (2G) systems use digital multiple access technology, such as TDMA (time division multiple access) and CDMA (code division multiple access). The Global System for Mobile Communications, or GSM, uses TDMA technology to support multiple users. Widely used second generation systems are GSM, which is the European standard.

There are several potential advantages of 2G systems over 1G:

1. 2G is digital technology based; and it increases radio system capacity and spectrum utilization efficiency. It also enhances voice quality due to the improved method of error correction mechanism.
2. 2G eliminates the major drawbacks of the 1G system by supporting roaming between network operators and between different countries. To do this, standards for core networks are introduced.
3. It supports not only circuit-switched voice communication, but also mobile data and Internet services.

A new design was introduced into the mobile switching center of second generation systems. In particular, the use of base station controllers (BSCs) reduces the load placed on the MSC found in first-generation systems. This design allows the interface between the MSC and BSC to be standardized. Hence, considerable attention was devoted for interoperability and standardization in second generation systems so that the carrier could employ different manufacturers for the MSCs and BSCs. Figure 2.4 shows the basic 2G network components.

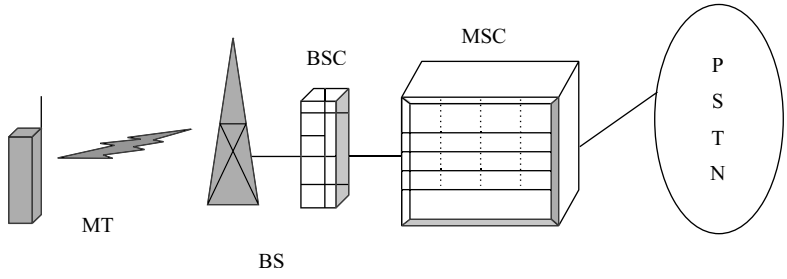


Fig. 2.4 *Basic 2G network components*

Major 2G Standards Figure 2.5 depicts the different network standards. The North American system IS-136 is based on TDMA technology, whereas IS-95 is based on CDMA technology. IS-136 is primarily used in the United States, whereas IS-95 is used in the United States and South Korea.

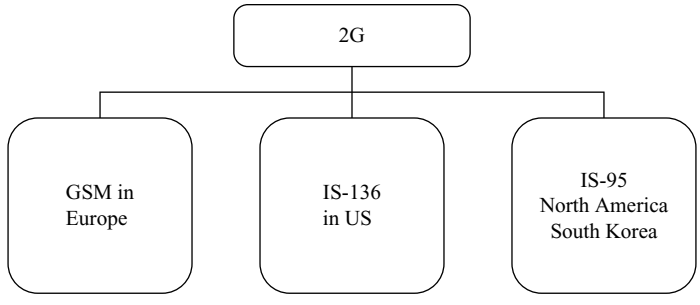


Fig. 2.5 *Different 2G network standards*

In Europe, GSM is the most widely used 2G network. European countries jointly developed a single set of RAN and the 2G core network replacing different 1G radio systems so as to support roaming between networks and between countries. GSM is used mainly for circuit-switched voice communication. The current GSM networks transmit data at 9.6 kbps with circuit-switched data transmission and allow up to eight users to share a single 200 kHz radio channel by allocating a unique time slot to each user. GSM is used in the 900 and 1800 MHz bands all over the world except for North America (1900 MHz band). The first wave of mobile data services was text-based instant messaging of up to 160 characters, and was succesfully introduced in Europe over GSM networks. SMS services grew very quickly over Europe and other countries.

In Japan, NTT DoCoMo has developed its own 2G radio system known as the Personal Digital Cellular (PDC) network. Like GSM, PDC supports circuit-switched voice and 9.6 kbps data communication.

Table 2.1 is a comparative chart that will give a quick overview of 2G wireless systems for different countries with their distinct features.

Table 2.1 *Comparative chart for 2G wireless networks*

Country	N. America	Europe	Japan	USA
Access network	IS-136	GSM	PDC	IS-95
Multiple accesses	TDMA/FDD	TDMA/FDD	TDMA/FDD	CDMA
Modulation	π/4 DQPSK	GMSK	π/4 DQPSK	QPSK/DQPSK
Downlink channel MHz (BS to MS)	869–894	935–960	810–826	869–849
Uplink channel MHz (MS to BS)	824–849	890–915	940–956	824–849
Channel spacing	30 kHz	200 kHz	25 kHz	1250 kHz
Transmission rate	48.6 kbps	270.833 kbps	42 kbps	1.2288 Mcps
Speech TX rate	7.95 kbps	13.4 kbps	6.7 kbps	1.2/2.4/4.8/9.6 kbps

The second wave of mobile data services is the low-speed mobile Internet service that was first succesfully launched by NTT DoCoMo over PDC networks in Japan in the year 1999. This service is known as *i-mode service*. It mainly supports email and instant messaging, and commercial transactions like banking, ticket reservation, credit card billing enquiry, etc. The major limitations of the i-mode service is low data rate and proprietory protocols (developed by NTT DoCoMo) instead of common IP-based access over the Internet.

2.5G Wireless Networks As the demand increases with the growing number of mobile users, data services, 2G circuit-switched-based data service with low rate (9.6 kbps) will not suffice. In time, 2G wireless networks have been enhanced to 2.5G with the introduction of new nodes in the core networks that support packet data service.

There are two wireless networks derived from GSM networks:

1. General Packet Radio Services (GPRS) that provide packet-switched core network as an extension to the existing GSM core network in order to support packet services over GSM radio systems.
2. Enhanced Data rates for Global GSM Evolution (EDGE) that uses advanced modulation and channel-coding techniques for increased data rates (384 kbps) over GSM.

2.5G world begun with the General Packet Radio Service (GPRS). GPRS is a radio technology for GSM networks that adds packet-switching protocols with shorter set-up time for Internet Service connections. Packet switching is a technique whereby the information (both voice or data) is sent into packets of a few kilobytes each at a time. These packets are routed using Internet protocol over the network between different destination addresses within each packet. Use of network resources is optimized compared to dedicated circuit-switched communication, as the resources are needed only during the handling of each packet. Billing for users is done by the amount of data sent, rather than connection time. GPRS system is characterized by

1. Different paths for voice and data transmission
2. Voice transmission—circuit-switched transmission
3. Data transmission—packet-switched transmission

The phase after GPRS is EDGE. EDGE is a radio-based high-speed mobile data standard that allows data transmission speed of 384 kbps, achieved using all eight timeslots. The main idea behind EDGE is to reach even higher data rates on the current 200 kHz GSM radio carrier, by changing the type of modulation used, but still working with current circuit switches. EDGE is sometimes known as enhanced GPRS system (EGPRS). Implementing EDGE will require relatively small changes to network hardware and software as it uses the same TDMA frame structure, logic channel and 200 kHz carrier bandwidth as today's GSM networks. As EDGE progresses to coexist with 3G WCDMA, data rates of up to ATM-like speeds of 2 Mbps could be available.

Due to the high data rate achieved by EDGE, it is sometimes considered as the next-generation 3G systems. EDGE will also be a significant contributor in 2.5G. It will allow GSM operators to use existing GSM radio bands to offer wireless multimedia IP-based services and applications at theoretical maximum speed of 384 kbps with a bit-rate of 48 kbps per timeslot. EDGE may provide similar kind of data services

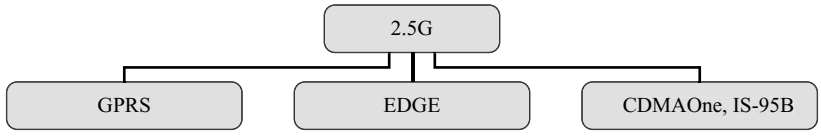


Fig. 2.6 *Three major 2.5G systems*

that a 3G user may get, but with the advantage that operators can function without 3G license. Fig. 2.7 is the 2.5G network architecture.

Like GPRS, CDMA based IS-95B is deployed worldwide providing high-speed packet and circuit-switched data access on a common CDMA radio channel. IS-95B simultaneously uses 8 different user Walsh codes to give 8 × 14.4 kbps throughput to a dedicated user. It also provides faster hard handoff.

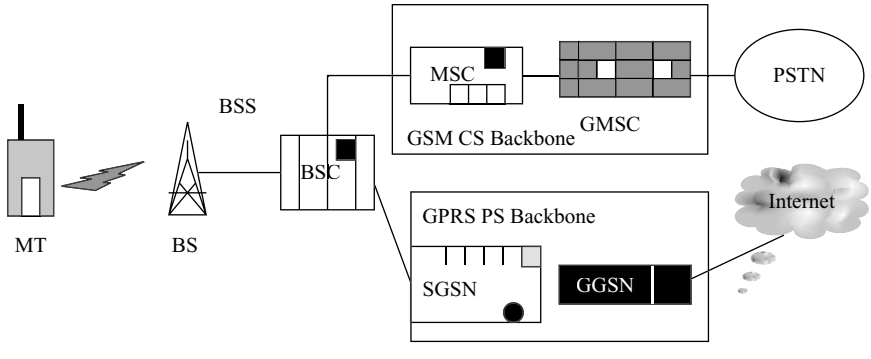


Fig. 2.7 *2.5G network architecture*

Another network known as High-Speed Circuit-Switched Data (HSCSD) is an enhancement of data services (circuit-switched data) of all current *GSM* networks. It allows users to access non-voice services at 3 times faster rate, which means subscribers are able to send and receive data from their portable computers at a speed of up to 28.8 kbps; this is to be upgraded in many networks to rates up to 42.2 kbps.

2.3.3 Third Generation (3G) Wireless Networks

Standardisation work for 3G wireless networks began in the late 1990s. The main perspective of 3G networks is to deliver high-rate voice and data service. The data rate for moving users is up to 144 kbps, 384 kbps for pedestrian speeds and up to 2 Mbps to stationary users.

The aim of 3G is to support IP-based data, voice and multimedia services with integration to Internet to provide useful Internet applications to mobile users. The improved interoperability to handle mobility across different radio technologies among different network providers is an important goal for 3G services.

It is to be mentioned that as 3G services aim to give real-time voice, streaming and non-real-time video, enhanced Quality of Service (QoS) is a prime factor for 3G networks. The 3G systems aims to provide multi-megabit Internet access with an 'always on' feature.

The International Telecommunication Union (ITU) formulated a plan (known as International Mobile Telephone, IMT 2000) to implement a global frequency band in the range of 2000 MHz to support single ubiquitous wireless communication all over the world. This ITU IMT 2000 standards organisation for 3G wireless systems is mainly grouped into two project groups—

The **UMTS** Group (3GPP) and the **CDMA2000** group (3GPP2)

Third Generation Partnership Project (3GPP) This is a partnership project for 3G mobile systems based on evolved GSM core networks and the radio access technologies that support GSM core networks. The radio access technology for 3GPP is known as Universal Terrestrial Radio Access Networks (UTRANs) based on Wideband CDMA (WCDMA) radio technology. The eventual 3G evolutions for GSM, IS-136 and PDC systems lead to WCDMA. The fundamental network part for WCDMA is GSM, as well as the converged version of GSM and IS-136 through EDGE. Figure 2.8 shows the UMTS evolution path from GSM.

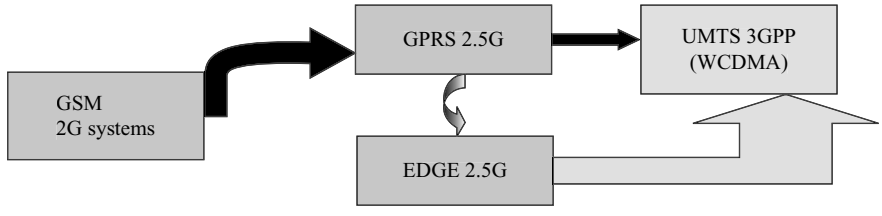


Fig. 2.8 *Evolution path to UMTS*

For 3G systems, the core network will comprise the GSM circuit-switched core and GPRS packet-switched core. The first release of specification of UMTS (Release 99) is focused to change the Radio Access Network, RAN, rather than the core network. This allows the core network to maintain its functionality, although changes will be made for the requirements of higher data rate for future networks. Mobile network operators continue with their existing infrastructure and will be in steps of 3G progresses. The handover between the UMTS and GSM is becoming one of the main criteria for 3G systems worldwide.

Third Generation Partnership Project 2 (3GPP2) It is a globally applicable standard for 3G having backward compatibility with IS-95. The access technology is based on CDMA2000.

IMT-2000 Process The IMT-2000 is a global process under ITU to develop next-generation mobile networks. It specifies the technical standards and allocates frequency. It is not a technology, but a system to allow seamless, ubiquitous user access to services to offer broadband real-time and non-real-time services.

Major IMT-2000 focus points for 3G are:

1. High-speed data transmission
2. Symmetric and asymmetric data transmission support
3. Improved voice quality comparable to fixed line network
4. Multiple simultaneous services to end-user
5. Support of global mobility between different operational environments
6. Improved security, capacity and spectral efficiency
7. Service flexibility

Table 2.2 *Data rate for 3G wireless networks*

User's condition	Data rate
Indoor pedestrian	2 Mbps
Mobile user (speed < 120 km/h)	384 kbps
Fast-moving vehicles	144 kbps

The process is intended to integrate many technologies under one roof under the IMT-2000 family. There are five different radio access technologies under IMT-2000 as given in Fig. 2.9.

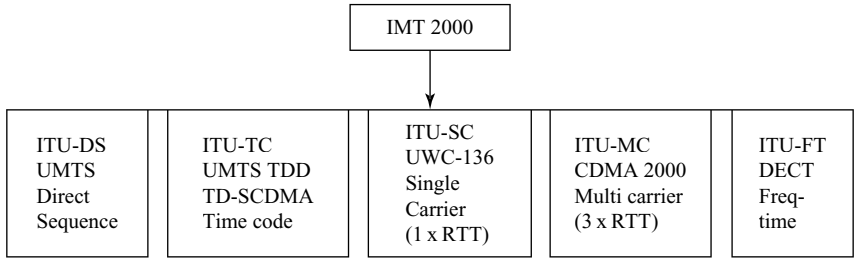


Fig. 2.9 *IMT 2000 family*

1. ITU-DS is the frequency division duplex (FDD) standard for UMTS
2. ITU-TC is the time division duplex (TDD) for UMTS and time division synchronous CDMA
3. ITU-MC is the multi carrier standard for CDMA2000
4. ITU-SC UWC-136 is the standard for EDGE
5. ITU–FT is the European standard for Digital Enhanced Cordless Telephones (DECT)

A key difference between the WCDMA and CDMA2000 is the multiple access technology they use. WCDMA uses two modes of direct sequence CDMA (DS-CDMA). One is FDD and the other is TDD. DS is the spread spectrum technology. It uses a PN sequence (pseudo random sequence) to spread the user traffic over the same frequency band. In the receiver site, the same PN sequence code is used to demodulate the data. Basically, FDD and TDD are the methods to separate the uplink and downlink traffic. For uplink data transmission, FDD uses 1920–1980 MHz, and for downlink traffic, it uses 2110–2170 MHz band. On the other hand, TDD uses different time schedules for uplink and downlink transmission over the same frequency band.

WCDMA requires a 5-MHz spectrum allocation. So the air interface for it needs a complete change for RF equipments at each base station. Due to the new set-up for a base station, implementation of 3G systems

becomes costly and slow. Through the evolutionary path of 3G, it is expected that by 2010–2015, WCDMA service will be fully installed, eliminating the backward compatibly with GSM/GPRS, IS-136, PDC and EDGE [3].

CDMA2000 uses FDD Muticarrier CDMA (MC-CDMA). A single-carrier CDMA2000 generally is known as 1 × RTT (Radio Transmission Technology). It provides instantaneous data rate to 307 kbps for a user in packet mode, but yields 144 kbps throughput per user. For a multi-carrier, 3 carriers (1.25 MHz each) are used together to achieve 384 Kbps data rate, and it is known as CDMA2000 3 × RTT. CDMA2000 wireless standards claim to give a much more seamless and less expensive upgrade path compared to WCDMA. It is because that CDMA2000 will also use the same frequency spectrum, bandwidth, RF equipments and air-interface framework at each base station when 3G upgrades would be introduced over time.

The ultimate objective for both 3GPP and 3GPP2 is to make an IP-based 3G core network through an evolutionary path, starting from wireless networks to full IP-based mobile networks [2].

Due to the higher data rate, 2.5G and the early version of 3G networks allow the user to access the Internet via IP-based protocols. Instead of voice and text-based instant messaging, the users can now use their mobile phones to take pictures and send them to another user, and send and receive email with multimedia contents like data and still pictures. Figure 2.10 shows the 3G evolution path.

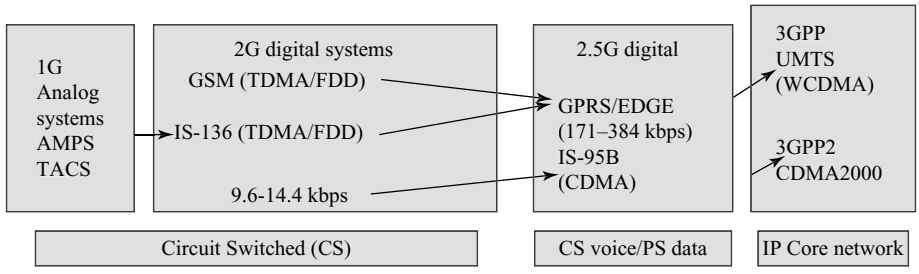


Fig. 2.10 *Evolutionary path for 3G*

2.4 WIRELESS LOCAL AREA NETWORKS (WLANs)

Wireless local area networks (WLANs) are of the IEEE 802.11 standard. It works similar to the traditional LAN except for the wireless interface. WLANs provide high-speed (11/54 Mbps) data communication in small areas such as a building or an office and also connect a user to IP networks, the Internet or enterprise networks.

WLAN uses the unlicensed ISM frequency band (Industrial, Scientific and Medical). The ISM band has three frequency ranges: 902–928, 2400–2483.5 and 5725–5850 MHz. The increasing number of PDAs and portable handheld devices enhance the popularity of WLAN, particularly in office and hot spot areas. It allows users to move around in a confined area while they are still connected to the network. The most popular widely used WLAN standard is the IEEE802.11b working at 2.4 GHz ISM band with a theoretical data rate of 11 Mbps. There are several other IEEE802.11 WLAN varieties that define physical layers (PHY) and the Medium Access Control (MAC) layer for WLAN with distinct purposes like security, QoS and increased data rate. Early development included industry-specific solutions and proprietary protocols, but at the end of the 1990s these were replaced by standards, primarily the various versions of IEEE 802.11 (Wi-Fi). Table 2.3 shows the existing WLAN family.

WLAN hardware was originally so expensive that it was only used as an alternative to cabled LAN in places where cabling was difficult or impossible, although the restricted range of the 802.11b (typically 30 ft) limits its use to smaller buildings. WLAN components are now cheap enough to be used in **home**, **enterprise** and **commercial public WLANs** (mainly in hot spot areas). Public WLANs are mainly used to provide mobile Internet services to business travelers and consumers.

WLAN uses the spread-spectrum technology based on radio waves for information transportation between devices within a limited coverage area of about 200 ft. There are two types of WLANs—infrastructure WLANs and independent Ad-hoc mode WLANs. Infrastructure WLANs, where the wireless network is linked to a wired network, are more commonly in use today. In an infrastructure WLAN, the wireless network is connected to a wired network such as the Ethernet, via access points, which possesses Ethernet links and antennas to send signals. These signals get circular coverage areas, depending on walls and other physical obstructions, in which devices can communicate with the access points.

The main transmission technology for WLAN is Spread Spectrum and Infrared. Frequency Hopping (FHSS) and Direct Sequence (DSSS) modulation are the two methods used by the spread-spectrum transmission.

Table 2.3 *WLAN family*

IEEE 802.11b	Defines a physical layer that provides data rates up to 11 Mbps in the 2.4 GHz ISM radio frequency band, most widely used WLAN today
IEEE 802.11a	Supports data rate up to 54 Mbps using the 5.7 GHz ISM radio frequency band
IEEE 802.11i	Defines a framework and means for supporting security over IEEE 802.11 WLANs
IEEE 802.11e	Defines a framework for supporting QoS for delay-sensitive applications (real time voice and video over IEEE 802.11 WLANs
IEEE 802.11f	Defines the Inter Access Point Protocol (IAPP) to assure interoperability of multi-vendor access points
IEEE 802.11n	The newest IEEE standard in the Wi-Fi category is *802.11n*. It was designed to improve on 802.11g in the amount of bandwidth supported by utilizing multiple wireless signals and antennas (called *MIMO* technology) instead of one. 802.11n connections should support data rates of over 100 Mbps.

1. **Frequency hopping:** The signal hops from one frequency to another within a given frequency range. The transmitter device listens to a channel, and if it detects an idle time it transmits the data using the full channel bandwidth. If the channel is full, the signal hops to another channel and repeats the process.
2. **Direct sequence modulation:** This method uses a wide frequency band together with Code Division Multiple Access (CDMA). Signals from different units are transmitted at a given frequency range. A code is transmitted with each signal so that the receiver can identify the appropriate signal transmitted by the sender unit.

The power levels of these signals are very low, just above background noise.

Infrared Transmission This method uses infrared light to carry information. There are three types of infrared transmission—diffused, directed and directed point-to-point.

1. **Diffused:** Infrared light transmitted by the sender unit fills the particular area, for example, an office. Therefore, the receiver unit located anywhere in that area can receive the signal.
2. **Directed:** Infrared light is focused before transmitting the signal. This method increases the transmission speed.
3. **Directed point-to-point:** Directed point-to-point infrared transmission provides the highest transmission speed. Here the receiver is aligned with the sender unit. The infrared light is then transmitted directly to the receiver.

The light source used in infrared transmission depends on the environment. A light-emitting diode (LED) is used in indoor areas, while LASERs are used in outdoor areas.

2.4.1 WLAN Architecture

The IEEE Project 802.11 specified two ways for implementing WLAN—ad-hoc mode and infrastructure mode. A number of mobile users meeting in a small region can set up an ad-hoc network. It does not need

any support from a wired/wireless backbone. In an infrastructure network, a cell is also known as a Basic Service Area (BSA). It contains a number of wireless stations. The size of a BSA depends on the power of the transmitter and receiver units. It also depends on the environment. A number of BSAs are connected to each other and to a distribution system by Access Points (APs). A group of stations belonging to an AP is called a Basic Service Set (BSS). Figure 2.11 shows the basic network architecture for wireless LANs.

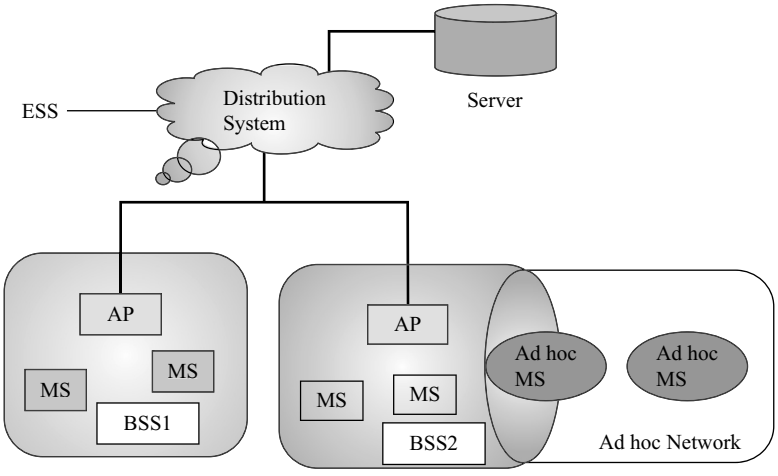
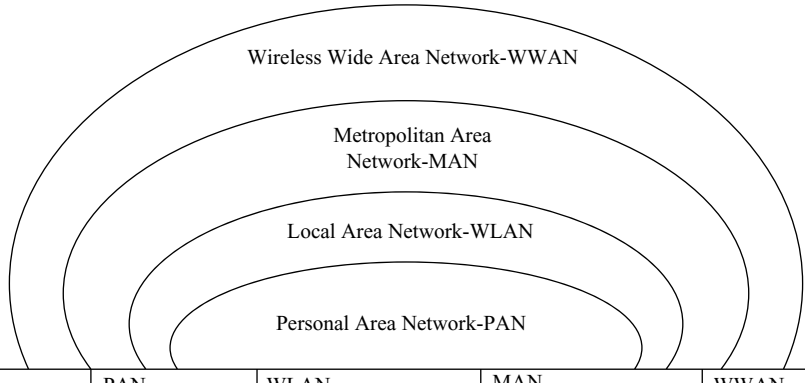


Fig. 2.11 *WLAN architecture*

2.5 WIRELESS TECHNOLOGY DIVISIONS

Figure 2.12 illustrates the different technology divisions at a glance using WLAN.



	PAN	WLAN	MAN	WWAN
Standards	Bluetooth	802.11a, 11b, 11g, Hiper LAN2	802.11, MMDS, LMDS, WiMax	GSM, GPRS CDMA, 2.5-3G, WiMax
Speed	< 1 Mbps	2 to 54 Mbps	22+ Mbps	9.6 to 384 kbps
Range	Short < 10 m	Medium < 100 m	Medium-Long < 1 km	Long
Applications	Peer-to-peer Device-to-Device	Enterprise Home networks	Fixed, Last mile access	PDAs, mobile phones, cellular access

Fig. 2.12 *Wireless technology divisions*

2.6 CELLULAR–WLAN INTEGRATION

Because of limited mobility and short transmission range, wireless LANs can be used in confined areas (meeting rooms, etc). The growing market of WLAN deployment in public and enterprise places will create special significance for mobile services with higher data rates that WLAN can provide. On the other hand, cellular communication networks provide wider coverage for high mobility users, but with a low data rate. These complementary characteristics of the two networks can be better utilised for public mobile services by seamlessly integrating the two technologies. Technology integration will provide adequate services to a user depending on the mobility and availability of the networks. Integration introduces new challenges for handover between two different systems, security and QoS. These issues are to be solved without changing the existing standards. Seamless handover will be required for users moving from one network to another ensuring the quality of service.

2.7 ALL-IP NETWORK: VISION FOR 4G

A rapid technological transition is occurring today in the world of internetworking. This transition is marked by the convergence of the communication infrastructure with that of IP data networking to provide integrated voice, video, and data services. As this transition progresses, the standards continue to evolve and many new standards are being developed to enable and accelerate this convergence of telecommunications and IP networking to mobilize the internet and to provide new multimedia services.

Mobile operators are transitioning towards 3G networks and beyond, i.e., 4G, in order to provide high-speed data access (100 Mbps) and multimedia services deploying several access technologies in a seamless converged IP-based network. Mobility management is the key issue for 4G mobile wireless networks. Some of the desirable design issues for 4G networks that require immediate attention are:

1. Smooth handoff both for intra/inter-domain network with reduced hand-off latency to support multimedia services
2. QoS support for multimedia application and real-time traffic management

'Beyond 3G' refers to an advanced level of 3G that uses the concept of an all-IP switching core. An all-IP switching core means that IP replaces the time division multiplex-based MSC with IP-based transport and IP-based signaling. IP-based signaling is implemented with new protocols like Session Initiation Protocol (SIP) and Media Gateway Control Protocol (MGCP). The traditional functionality of MSC is modified for All-IP switching, and the various MSC functions are redistributed to several other elements. A good example of this evolution in the switching core from TDM to packets is 3GPP's R4 and R5 architectures. 3GPP2 also has adopted a similar trend to transition to an all-IP network. An initiative is also underway to develop end-to-end all-IP mobile wireless networks in which both circuit and packet switch core and the radio access network are IP-based.

Again, 3G network is yet to be deployed widely. The service cost of 3G will be greater for future requirements and applications because of the high cost of deployment. Thus, the next generation, or fourth generation (4G), networks are coming into the picture. 4G will be based on all-IP concepts, mainly focusing on wireless IP with self-provisioning of different multimedia services like video, audio and games. The fundamental reason for all-IP concepts is to have a common platform for all heterogeneous technologies that have been developed so far.

1. IP is compatible with and independent of the actual radio access technology.
2. With IP, one basically gets rid of the lock existing between the core networking protocol and the link layer, the radio protocol.
3. IP tolerates a variety of radio protocols. One could be a core network provider that supports different access technologies like 802.11, WCDMA, etc.
4. All-IP networks may provide greater mobility with lower costs.

The 4G standards have not yet been set and underlying technologies are not fixed, which leaves plenty of opportunities for other applications to overlap into the 4G space. The word **'MAGIC'** is just appropriate to define 4G. It signifies **M**obile multimedia, **A**nytime anywhere access, **G**lobal mobility support, **I**ntegrated wireless solution and **C**ustomized personal service. We are totally connected through better telephone lines—wired or wireless—the Internet, wireless and gadgets that allow us to stay connected anytime and anywhere. The technological development is happening so fast that a new gadget replaces an older one even, before the latter has captured the market. The same thing happens in case of WiFi and WiMax—a technology for broadband wide area networks.

The goal of 4G is to replace the entire core of cellular networks with a single worldwide cellular network completely standardized and based on the IPv6 for video (Internet Protocol), packet-data utilizing, Voice-over IP (VoIP) and multimedia services. The newly standardized networks would provide uniform video, voice, and data services to the cellular handset or handheld Internet appliance, based entirely on IP (Internet Protocol)—perhaps satisfing the name 4G cellular.

4G is characterized by heterogeneity in architecture, protocols and air interfaces. Therefore, it is compatible with all common network technologies like 802.11 (WLAN), 802.16e (Mobile WiMAX), 802.20, W CDMA, Bluetooth and HyperLAN. Wi-Fi and WiMAX 3G, with other wireless technologies, are under development to meet the standards of performance for 4G deployments. The face-off between 4G verses Wi-Fi verses WiMAX will be a globally adopted network of shared applications, resources and standards satisfying a minimum delivery requirement for VoIP, broadband, and multimedia services to all 4G enabled end users.

2.7.1 Key Technologies for 4G

One of the key 4G technologies is OFDM—Orthogonal Frequency Division Multiplexing. Use of this scheme in the physical layer not only provides advantages, it also helps L2 performance. OFDM exploits time, space, frequency and code domain to optimise radio channel performance, takes care to effect of multi-path propagation and reduces receiver design complexity. OFDM is also compatible with MIMO—Multiple Input Multiple Output technology and smart antenna. The two together enhance system throughput.

As the data rate and bandwidth requirements for 4G are much higher (20–100 Mbps, 100 MHz) than 3G (384 kbps–2 Mbps, 5 MHz), the key technologies for 4G will be OFDM and muti-carrier CDMA (MC-CDMA) for the physical interface. The fundamental difference between 2G, 3G and all-IP is that the functionality of the radio access network and base station controller is distributed to the BTS (Base Transmission System) and a set of server and gateways. This leads cheaper network access and will provide faster operations. Figure 2.13 shows the conceptual architecture for an all-IP network.

IP-BTS functions partly as RNC/BSC, and is not capable of performing layer 1, 2 and 3 functions. IP servers handle the signal between the network elements at the layer 3 and tune network parameters for efficient radio resource managements. As there are multiple access technologies, a common server can improve the efficiency of the networks instead of having multiple servers for different access radio interfaces. Gateways are responsible in interacting between IP-based RAN and IP-based core networks. It can be of two types—circuit-switched gateway and packet-switched gateway.

4G is yet to be standarised. The path for 4G may come in two ways—evolutionary and revolutionary. It is expected that 4G will be just like an umbrella—the convergence platform of heterogeneous networks. 4G will be the fully IP-based integrated system of systems and network of networks achieved after convergence of wired and wireless networks [7].

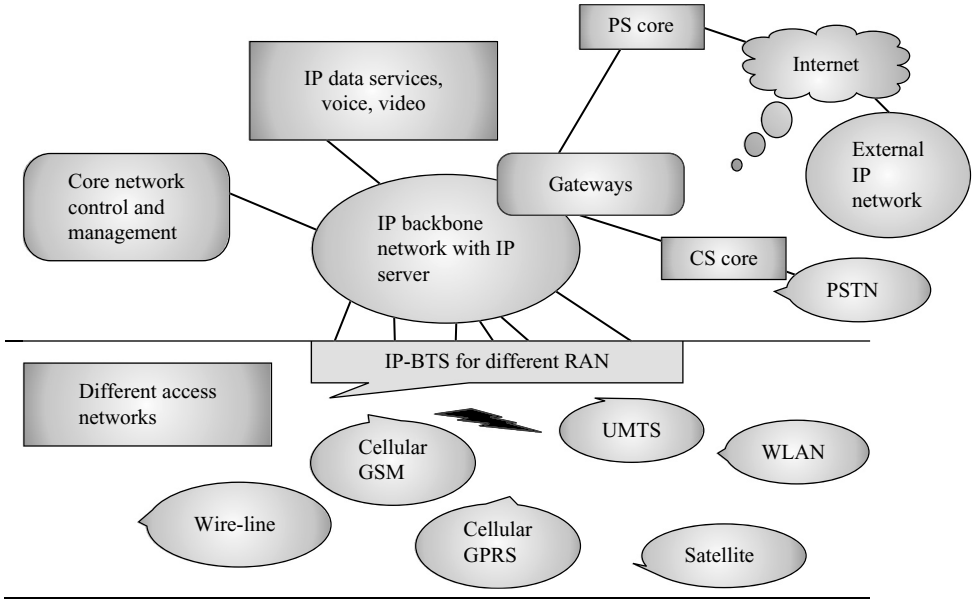


Fig. 2.13 *Example of All-IP network architecture*

Challenges are many in the way of realizing 4G dreams. Consolidated solutions that can seamlessly operate on the multiple, diverse networks migrating to the 4G environments obviously invite new challenges on every step and researchers worldwide face an uphill task of designing suitable solutions for the following:

1. Automatic network tracking
2. Multi-standard/Multimode user terminals
3. Multi-operator-oriented intelligent billing system, packet-switched oriented billing, QoS dependant charging, real-time billing
4. Personal and session mobility—confusions regarding the choice of either MIP or SIP as the core protocol
5. Integration and interoperability of diverse networks, QoS maintenance-varying bit rates, bandwidth allocation, channel characteristics, fault-tolerance levels and hand-off management

There is still a long way to go before a universally accepted, completely transparent, user friendly, cooperative public–private wireless broadband communication framework comes into play.

Summary

The generation of mobile communication networks have traversed a long way through different phases of evolution since its birth early in the 1970s. The steady global boom in the number of mobile users each year has periodically spurred the development of more and more sophisticated technologies trying to strike the right chord primarily in terms of provision of seamless global roaming, quality services and high data rate. Today numerous different-generation technologies with their individual pros and cons exist globally. The coming era of 4G systems foresees a potential smooth merger of all these heterogeneous technologies with a natural progression to support seamless cost-effective high data rate global roaming, efficient personalised services, typical user-centric integrated service model, high QoS and overall stable system performance. However, every step in such technological advancements presents huge research challenges.

References

[1] Abu El-Ata M.A, *Evolution of Mobile Cellular Communication Systems: The Journey to UMTS*, 17th National Radio Science Conference, NRSC'2000.

[2] Cheng J.C., and T. Zhang, *IP-Based Next Generation Wireless Networks–Systems, Architecture and Protocols*, John Wiley and Sons, INC, 2004.

[3] Rappaport T.S., *Wireless Communications–Principles and Practice*, Pearson Education India, Second edition, 2003.

[4] Bannister J.P., Mather and S. Coope, *Convergence Technologies for 3G Networks–IP, UMTS*, EGPRS and ATM, John Wiley and Sons, Ltd., First Edition, 2004.

[5] Jaseemuddin M., *Architecture for Integrating UMTS and 802.11 WLAN Networks,* IEEE Symposium on Computers and Communications-ISCC, Turkey, pp.1–8, 2003.

[6] Hui S.Y., and K. H. Yeung, *Challenges in the Migration to 4G Mobile Systems*, IEEE Commuications, Vol. 41, No. 12, pp. 54–59, Dec. 2003.

[7] Weiss Kim, *Strategic Options for Managing Technology Evolution in Wireless Industry*, White paper, March 2004.

[8] Kim Y.K., and R. Prasad, *4G Roadmap and Emerging Communication Technologies*, Artech House, Boston–London, 2006.

[9] *IEEE 802.11 WG, Part 15:Wireless Medium Access Control and Physical Layer Specifications for High Rate Wireless Personal Area Networks (WPAN)*, D14, Oct 2003.

[10] *IMT-2000: Standards Efforts of the ITU*, IEEE Personal Communications, Vol. 4, No. 4, August 1997.

[11] http://www.gsmdata.com/cannes99/cannespaper.htm

Questions for Self-Test

2.1 The most widely used standard for cellular communications is
- a. the advanced mobile phone service (AMPS)
- b. the mobile subscriber unit (MSU)
- c. the mobile telephone switching office (MTSO)
- d. code division multiple access (CDMA)

2.2 How many conversations per channel can TDMA digital cellular carry at once?
a. 1 b. 2 c. 8 d. 10

2.3 Which of the following is not a limitation of AMPS?
a. Low calling capacity b. Poor privacy protection
c. Limited spectrum d. Wide coverage area

2.4 Digital cellular technologies offer increased capacity and security.
a. True b. False

2.5 TDMA, a digital air interface standard, has twice the capacity of analog.
a. True b. False

2.6 Both AMPS and TACS use frequency modulation technique for radio transmission.
a. False b. True

2.7 The GSM frequency band for North America is
a. 1900 MHZ b. 1800 MHz

2.8 Billing for GPRS users is done by the amount of data sent, rather than connection time.
a. False b. True

2.9 A key difference between the WCDMA and CDMA2000 is the multiple access technology they use.
a. True b. False

2.10 WCDMA uses two modes of direct sequence CDMA (DS-CDMA). One is FDD and the other is TDD.
a. True b. False

2.11 CDMA2000 uses FDD Muticarrier CDMA (MC-CDMA).
a. False b. True

2.12 HSCSD is the enhancement of GSM system with higher data rate.
a. True b. False

2.13 What are the main characteristics of AMPS networks?

2.14 List different 1G systems available with their main features.

2.15 What are the limitations of 1G cellular networks?

2.16 Digital technology is the breakthrough for 2G cellular systems—explain.

2.17 What are the two main 2G systems evolving in American and European countries?

2.18 Discuss the main points in the evolutional process of 1G to 3G.

2.19 Depending on the coverage areas, how many different wireless networks are defined? Discuss the main features of each network.

2.20 DECT system was mainly used for domestic and small business organization—explain. What were the areas of application for the DECT system?

2.21 Compare and contrast the 2.5G GPRS and EDGE systems.

2.22 EDGE is sometimes considered as 3G system—explain.

2.23 Make a table for different available wireless systems for 1G to 3G.

2.24 Discuss the main objectives of IMT-2000. What are the two main project branches under IMT-2000?

2.25 What are the main objectives of 3G systems? Describe the evolutionary path track for 3G systems with a pictorial representation.

2.26 WLAN is a low-cost data service for limited coverage—explain.

2.27 What are the main transmission technologies for WLAN?

2.28 Discuss the necessity of cellular-WLAN integration.

2.29 What is the driving force for 4G technologies? What basic changes are required for implementing 4G from other 1G to 3G networks?

2.30 Discuss the fundamental concepts of all-IP networks.

Cellular Mobile Wireless Networks: Systems and Design Fundamentals

3

Introduction

The radio spectrum bandwidth is scarce. Today's boom in wireless communication limits the maximum number of user's capacity that can be supported in a wireless system and thus becomes an important parameter for performance evaluation of a network. In cellular wireless communication, a cell is defined as the radio coverage area by a single base station. If there is a single base station, then it will require a high power transmitter to support users for wider coverage. With a single base station and single large transmitting antenna, the system capacity will hit the limit. The system capacity can be considered as the maximum number of users that can be supported within a channel. To support larger user base, a larger geographical coverage is required. To use radio resource very efficiently, unique arrangement of cellular array with low power transmitter to each cell is the fundamental concept of cellular wireless communication. The system capacity can be enhanced by frequency reuse technique. The concept of frequency reuse plays a good role to serve smaller geographical area by a single cell and is used in another cell in such a manner to avoid interference between the users. This calls for the radio frequencies to be physically separated by each other. The power level of the signal from one cell to the other to use the same frequency must be within an acceptable limit of interference. This process of replicating small identical geographical structure over a larger area gives rise to the concept of cellular communications. Cellular communication also called PCS, Personal Communication Services aims to reach people anywhere and anytime to provide telephone services.

A more efficient use of limited wireless resources in personal communication services requires much smaller cells, micro or pico cells than the conventional cellular networks. The increase in system capacity may be achieved by the use of smaller cells, frequency reuse pattern and use of antenna within a sector. As the cell size decreases, the interference effects between the adjacent cells and co-channel cells come into play. Interference on voice channels causes cross talk, where the subscriber hears interference in the background noise. This interference effect becomes more prominent in the urban areas because of more radio noises due to large number of base stations and mobile devices. So, signal-to-interference ratio is the design-determining factor and is the cause of cell planning.

Mobility Management is the new paradigm in wireless communications. It is again divided into two parts: Location Management and Hand-off Management. Use of efficient algorithms for both; help to provide seamless communication among the users while roaming within wireless networks across the cells. Radio resource in wireless communication is the most precious. So, Radio Resource Management plays an important aspect of wireless communication for providing better quality of service and access to more and more number of users. Performance of wireless communication networks depends on the tele-traffic modeling, call admission control mechanism including the way of hand-off management.

In this chapter, all the important aspects for the design of wireless communication have been discussed with the above consideration pointed out.

3.1 DESCRIPTION OF CELLULAR SYSTEM

The design of a wireless cellular system considers the cell-dividing concept. The total geographical coverage area is divided into smaller areas. Thus the wireless network allows the use of smaller power transmitters and efficient use of radio spectrum by means of frequency reuse. Prior to the introduction of cellular radio, mobile

wireless cellular service was only provided by a high power base trans-receiving system. A typical system would support about 25 channels and with an effective area of 80 km. The cellular system is limited by interference than the conventional system with large single base station that is limited by thermal noise. In this section, the cellular structure and related parameters are discussed to describe the design of cellular system. The word "Mobile" is a fast moving radio terminal (handset) having wireless network access, such as the cellular telephone system. In a radio cellular communication – a radio cell is defined as the geographical area served by a base station through which mobile users can communicate. Variable power levels allow cells to be sized according to the subscriber density and demand within a particular region. To increase capacity, cellular wireless systems use frequency reuse techniques. Channels used in one cell can be reused in another cell located at some distance. Cells can be added to accommodate growth, creating new cells in unserved areas or overlaying cells in existing areas.

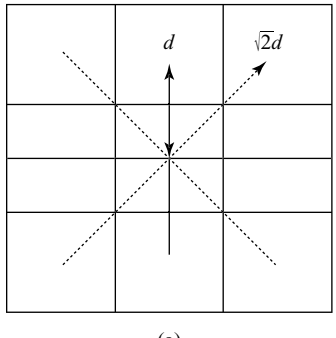
3.1.1 Cellular Structure

In principle, a given geographical coverage area is divided into subareas, each known as a *cell.* A cell is an area of coverage under a single base station within which signal reception conforms to the system specification. The essence of a cellular network is the use of multiple low power transmitters of the order of 100 Watt or less. A cell is assigned a band of frequencies and is served by a base station consisting of Trans-receiving (Tx-Rx) system and control unit.

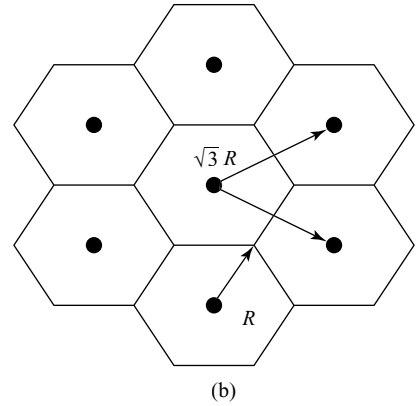
Extension of the coverage area of a base station depends on several factors. These are base transmitting output power, frequency band of operation, antenna height and location, antenna type, terrain geography and receiver sensitivity. Radio wave propagates from a base station in a straight line, i.e., line-of-sight propagation. So, a mobile station located behind a large obstacle or a tunnel or hill may not receive signals because of the out-of-radio coverage called the area of shade.

Shape of the Cell The first design decision was to make the shape of the cell such as to cover an area. If the shape is of uniform square area, then a cell has four neighboring cells at a distance d and four cells at a distance $\sqrt{2}d$ (Fig. 3.1a). When a mobile user moves across the cell boundaries he should get the signals equally from equidistant antennas in the neighboring cells. This is not happening in case of square cells.

So, ideally hexagonal cell patterns are considered that provide equidistant antennas to the neighbor cell sites. For a cell of radius R, the distance between the cell center and each adjacent cell center is $d = \sqrt{3}R$. This is shown in Fig. 3.1b.



(a)



(b)

Fig. 3. 1 *(a) Square shape cell, (b) Hexagonal cell structure*

Example 3.1 In a cellular communication system, in addition to the hexagonal topology, a square or an equilateral triangle topology can also be used. It is given that the distance between the cell center and its furthest perimeter points is R.

(a) Compare the cell coverage areas among the three regular polygons (hexagon, square and triangle).

(b) Discuss the advantages of using the hexagonal cell shape over the square and triangle cell shapes.

Solution

(a) We have to compare the coverage areas of these three different cells.

Let the length of each side of the square be s units.

So,

$$\frac{s}{\sqrt{2}} = R, \text{ where } R \text{ is the common radius in this case.}$$

So, $s = \sqrt{2}R$

Area $s^2 = 2R^2$ (i)

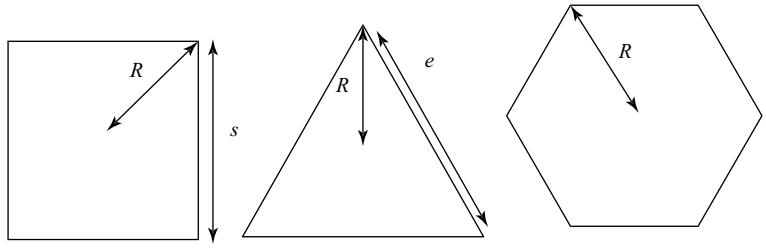


Fig. 3.2 *Different shapes of cells*

For an equilateral triangle, let the length of each side be e units.

Now,

$$\frac{2}{3} \times \frac{\sqrt{3}}{2} \times c = R$$

or,
$$e = \sqrt{3}R.$$

Hence, area covered by the equilateral triangle is

$$\frac{\sqrt{3}}{4} e^2 = \frac{\sqrt{3}}{4} \times 3R^2$$

$$3.\frac{\sqrt{3}}{4} R^2 \quad \text{(ii)}$$

$$= 1.3\ R^2 \text{ (approx.)}$$

For the case of a hexagonal cell, we divide it into six equal equilateral triangles.

Area of each equilateral is $\frac{\sqrt{3}}{4} R^2$

Total area covered by the hexagon is $\left(\frac{\sqrt{3}}{4} \times R^2 \times 6 \right) = 2.6R^2 \text{(approx.)}$ (iii)

Comparing (i), (ii) and (iii), it is evident that the cell coverage area is greatest for hexagon.

(b) Advantages of using the hexagonal cell shape over the square and triangular cell shapes
The hexagonal cellular structure has been universally accepted because its mathematical aspects are sound. The distance of the corners from the center of each cells is same, unlike in square cells. Thus the equal power will be delivered to each mobile residing at the corner of the hexagonal cell from the centrally placed base station.

While it might seem natural to choose a circle to represent the coverage area of a base station, adjacent circles cannot be overlaid upon a map without leaving gaps or creating over lapping regions. Thus, when considering geometric shapes which cover an entire region without overlap and with equal area, there are three sensible choices – a square, an equilateral triangle, and a hexagon. A cell must be designed to serve the weakest mobiles within the footprint, and these are typically located at the edge of the cell. For a given distance between the center of a polygon and its farthest perimeter points, the hexagon has the largest area among the three. Thus by using the hexagonal geometry, the fewest number of cells can cover a large geographic region, and the hexagon closely approximates a circular radiation pattern which would occur for an omni-directional base station antenna and free space propagation.

Omni-directional Cell In this category of cells, the base station antenna is of omni-directional type, i.e., an antenna that transmits equal power in all directions, in the azimuthal plane forming an approximate circular coverage area. At the center a base station is situated itself.

Sector Cell In a sector cells, the base station in each cell is equipped with a directive antenna to cover a specific region. Cells can be divided into several sectors defined with the fixed angular distribution.

3.1.2 Cell Cluster

A group of cells with different frequency is known as *cluster*. The number of cells forming cluster is differentiated according to the cell structure. There are several different standards for clustering as given below.

1. 4 cell standard with all omni-directional cells.
2. 7 cell standard with all omni-directional cells.
3. 12 cell standard with all omni-directrional cells.
4. 21 cell standard with seven base station, each base station associated with three sector cells.
5. 24 cell standard with four base stations, each base station associated to three sector cells.

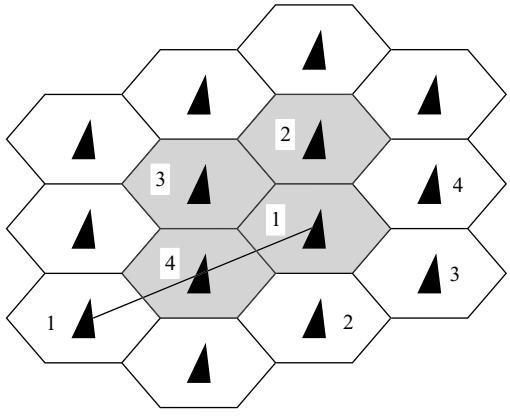


Fig. 3.3 *Set of four omni-directional cells (i=2, j=0).*

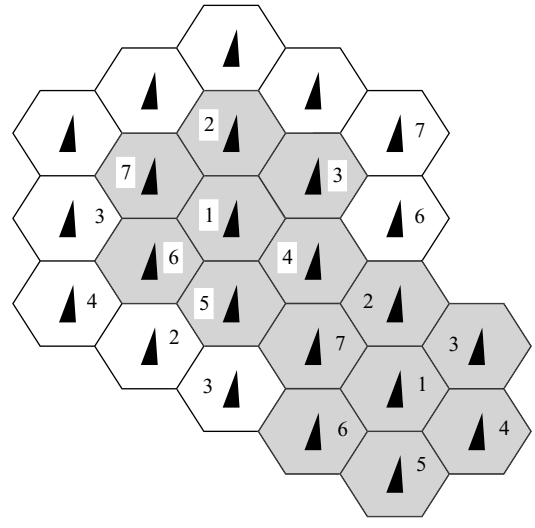


Fig. 3.4 *Set of omni-directional 7 cells cluster (i=1, j=2)*

In a cluster with seven cells, seven different frequencies A, B,…G can be used. But in a cluster with 21 sector cells, a more convenient designation A1, A2, A3, B1, B2, B3, ……, G1,G2,G3 are used. Figures 3.3 to 3.7 represent different cells clustering configurations. We shall describe the meaning of the parameters *i* and *j* later.

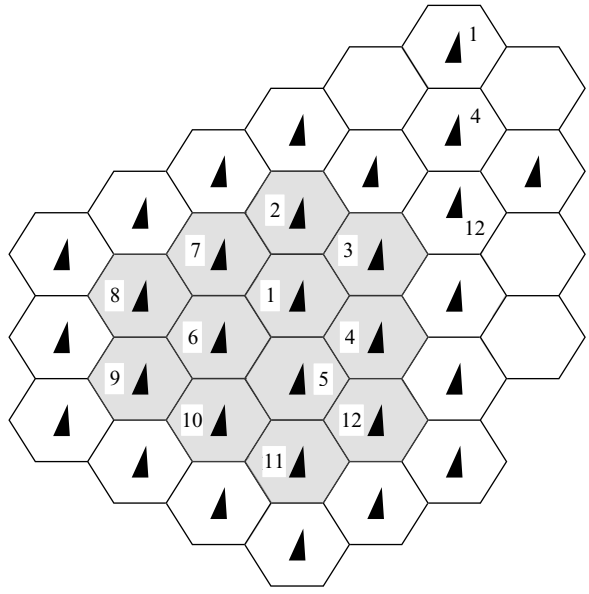


Fig. 3.5 *Set of omni-directional 12 cells cluster (i=2, j=2).*

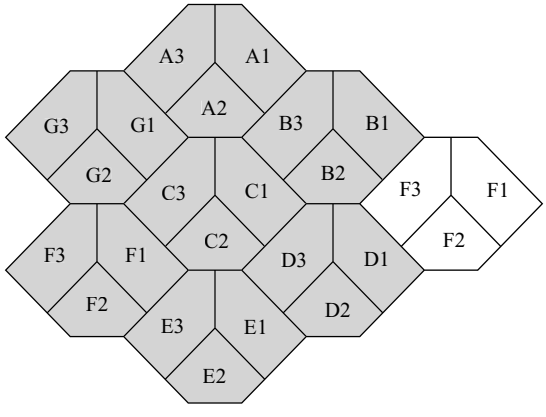


Fig. 3.6 *Set of sector cells using frequency reuse factor N=7 and 3 sectors*

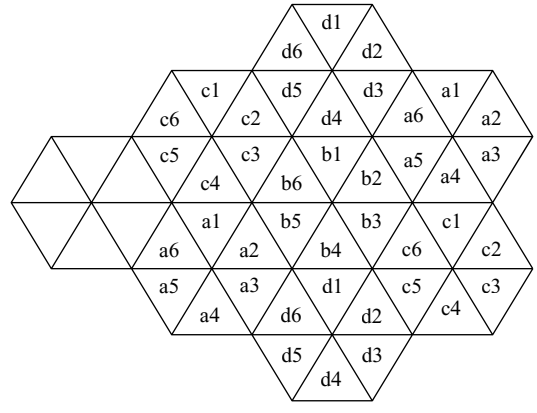


Fig. 3.7 *Set of four cells with 6 sectors*

3.1.3 Frequency Reuse

If a given set of frequencies or radio channels can be reused without increasing the interference, then large geographical areas covered by a single base station with high transmitting power antennas can be divided into small areas each allocated with a subset of frequencies. With a smaller geographical coverage, low power transmitter antenna is used. If the physical separation of the two cells is sufficiently high so that same subset can be used in both cells, then this concept is called **frequency reuse**. Frequency reuse means simultaneous use of same frequency in different distinct sets of cells. The distance between the cells of same frequency is limited by the maximum **co-channel interference** allowed in the system.

By using frequency reuse, the system capacity can be increased without the use of high power transmitter. But this advantage does not come without price. As the geographical areas are divided into small parts with cellular structure served by a single base station, this system restricts the user movement pattern. As mobile user moves from one cell to others, it requires hand-off to continue the ongoing services without disruption and maintaining end-to-end Quality of Service (QoS).

Two or more different cells can use the same set of frequencies or radio channels (co-channel cells) if the nearest cells are separated such that the interference between cells at any given frequency is at an acceptable level. The adjacent cells having same frequency are known as **co-channel cells**. The space between the adjacent co-channels is filled with cells having different frequency to maintain frequency isolation. The group of cells that uses different sets of frequencies is the cell cluster. If there are **N** number of cells within a cluster and **C** be the total number of available channels without frequency reuse, then N cells in each cluster can use all the C channels, i.e, each cell uses 1/N th channel, so, N is called the frequency reuse factor.

Determination of Frequency Reuse Factor N For homogeneous system, N is given as

$$N = i^2 + j^2 + ij \tag{3.1}$$

where, *i* and *j* are the positive integers and measure the number of nearest neighbors between the co-channel cells. For *i* =2 and *j* =1, N =7, the cell cluster consists of 7 cells (Fig. 3.3). The cellular layout with the concept *i* and *j* can be produced with the rules given below:

Move *i* cells along the hexagonal chain and then take 60 degree anti-clockwise turn and move j cells. This is illustrated in Fig. 3.8.

Possible values of N are1,3,4,7,9,12,13,….. In characterizing frequency reuse, the important parameters are:

D = the minimum distance between centers of cells that use the same frequency band called co-channel,
R = radius of hexagonal cell,
d = distance between the centers of adjacent cells and
N = the number of cells in a cluster that repeats, known as frequency reuse factor.

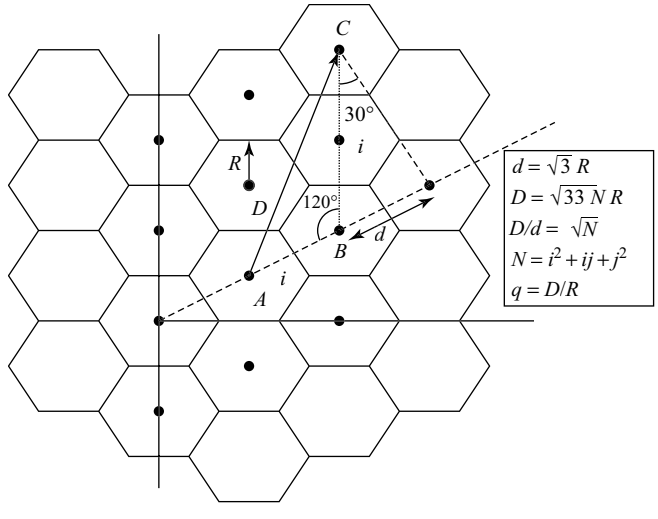


Fig. 3.8 *Geometry to calculate nearest co-channel cell distance i = 1, j =2*

The reuse distance D is given by the expression

$$D = \sqrt{3N}\ R \Rightarrow (D/d) = \sqrt{N}$$

where, R is the cell radius. This can be proved very easily with the help of Fig. 3.7.

Using the law of cosines for ΔABC,

$$\text{Cos} \angle \text{ABC} = \text{AB}^2 + \text{BC}^2 - \text{AC}^2 - 2.\,\text{AB.}\,\text{BC}$$

$$\text{Cos}\,120° = (i^2d^2 + j^2d^2 - \text{D}^2) / 2ijd^2$$

$$\text{D}^2 = d^2(i^2 + j^2) + ijd^2$$

$$(\text{D}/d^2) = (i^2 + j^2) + ij = \text{N}$$

From the Fig. 3.7, it can be proved in other way as

$$(\text{D}/d^2) = j^2 \cos^2 30° + (i + j \sin 30°)^2$$

$$= i^2 + j^2 + ij = \text{N}$$

So, $D/d = \sqrt{N}$ and $D/R = \sqrt{3}N = q$, called the normalized frequency reuse ratio. Figure 3.9 shows the relationship between D and R for first tier co-channel interfering cells. As N increases, 'q' also increases. Smaller value of N means to increase the capacity of the cellular system with the increase of co-channel interference due to the nearness of the co-channel cells. So, trade-off is required to select *q* and N such that signal to co-channel interference ratio is within tolerable limit. For particular values of *i* and *j*, N can be obtained from which '*q*' is calculated as given in Table 3.1.

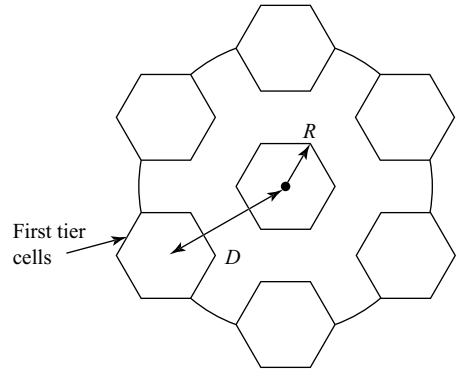


Fig. 3.9 *Relation between D and R*

Table 3.1 *Relation between q and N*

Frequency Reuse factor N	Corresponding value of (*i*, *j*)	Frequency Reuse Ratio q =D/R
3	1,1	3.0
4	2,0	3.46
7	1,2	4.58
9	3,0	5.2
12	2,2	6.0

It is clear from Table 3.1 that as the size of the cluster decreases, quality of transmission degrades because of possible interference effects, but the capacity increases due to possibility of distributing all channels among a few cells. As the D/R ratio increases, channels/cell decreases and traffic capacity goes from higher end to lower end, but transmission quality goes the reverse trend.

As D/R increases → Channel/cell decreases and Traffic capacity → High to Low and Transmission quality → Low to High.

Thus, a small value of '*q*' provides larger capacity since the cluster size N is small, whereas the larger value of *q* improves transmission quality due to smaller level of co-channel interference.

Interference is the main factor to determine the size of the cluster. There must be substantial distance between two adjacent co-channels for acceptable interference level.

3.1.4 Co-channel and Adjacent Channel Interference

For hexagonal cells, there are six nearest co-channel neighbors to each cell (Fig. 3.10). Co-channel cells are placed in tier so that 6*k* cells in the k^{th} tier surround a candidate cell. As D is the distance between the two adjacent co-channel cells, the radius in the k^{th} co-channel tier is $\sqrt{k}$ D for $k = 1$, and $\sqrt{k}$ +1 D for $k \geq 2$.

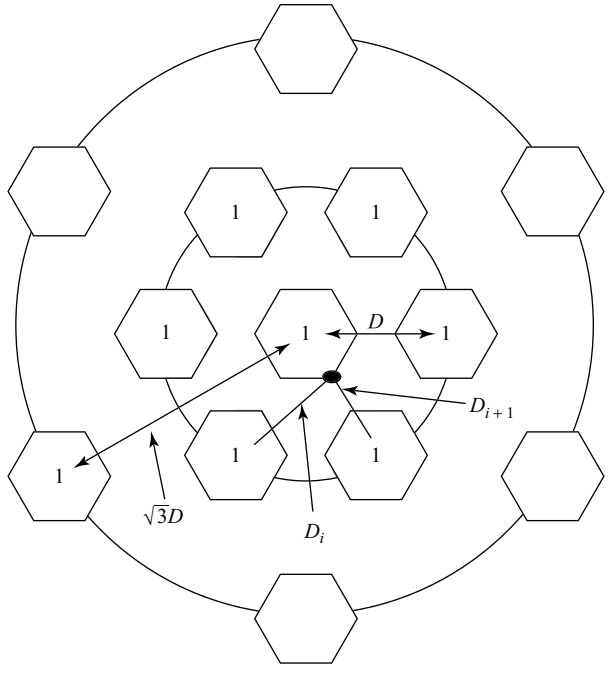


Fig. 3.10 *Co-channel cell tiers*

A given base station provides services to many mobile users. The signal received by a base station is subject to interference by other mobile transmission within the same cell site and also due to the transmission from other mobiles belonging to the neighboring cells. Interference from the other mobiles residing at the same cell is known as **intra-cell interference**, whereas interference between the different cells is known as **inter-cell interference**. To nullify the effect of inter-cell interference for downlink transmission, mobile handset receiver has to be sophisticated. But in general, uplink transmission is not very interference problematic because of the advanced cell-site receiver. Though the frequency reuse enhances system capacity, at the same time introduces co-channel interference.

Co-channel Interference Inter-cell interference from different cells is dominated by co-channel interference. So, any wireless communication system performance is to be characterized with the effect of interference from the co-channel cells. This phenomenon is complex if we consider the multi-path effects and propagation channel characteristic due to shadowing. At first if we consider only the effect of distance dependent path loss, we can simplify the signal to interference ratio. Let S is the signal power and I is the co-channel interference at the output of the receiver demodulator.
The signal to co-channel interference ratio at the desired mobile receiver is given by

$$\frac{S}{I} = \frac{S}{\sum_{i=1}^{N_I} I_i} \tag{3.2}$$

where, N_I is the number of co-channel interfering cells and I_i be the interference power caused by transmission from the i^{th} co-channel cell base station. The distance power law between the transmitter and the receiver decreases the average received signal strength. If the distance between the i^{th} interferer and the mobile unit is D_i and κ is the path loss exponent (varies from 2 to 5), then the received interference I_i is

proportional to $(D_i)^{-\kappa}$. On the other hand, the received signal power S is proportional to $(R_{mbs})^{-k}$, where R_{mbs} is the distance between the mobile and the serving base station.
By substitution we can write

$$\frac{S}{I} = \frac{(R_{mbs})^{-\kappa}}{\sum_{i=1}^{N_l} D_i^{-\kappa}} \tag{3.3}$$

The worst case of co-channel interference is occurred when the mobile is located at the cell boundary (i.e., $R_{mbs} = R_{cell\text{-}radius}$). Considering the hexagonal cell structure and co-channel interference only from the first tier cells, neglecting the higher tier cells, the number of nearest neighboring co-channel cells $N_I = 6$. For the mobile located at the boundary of the cell and with $D = D_i$, i =1, 2, 3, …6, as shown in the Fig. 3.9,

$$\frac{S}{I} = \frac{(D/R_{mbs})^{\kappa}}{N_I} = \frac{q^{\kappa}}{N_I} = \frac{\left(\sqrt{3N}\right)^{\kappa}}{N_I} \tag{3.4}$$

$N_I = 6$ in our consideration.
From this, frequency reuse ratio (q) can be easily determined as, $q = (6 \text{ x } S/I)^{1/\kappa}$.
Let us consider an example problem for the explanation of co-channel interference effect.

Example 3.2 A cellular telephone system with 110 channels uses a modulation scheme requiring a minimum S/I ratio of 19 dB for acceptable link performance. Assume that the propagation loss is only distance dependent and increases with the fourth power of the distance. Determine how many channels/ cell can be offered by the system. Assume the hexagonal cell structure with base station at the center transmitting same power.

Solution For hexagonal cell plan, the neighbor co-channel cells are 6 at a distance D as shown in Fig. 3.9. Six additional cell at a distance $\sqrt{3}$D, 6 at a distance $\sqrt{4}$D …….and 6 at $\sqrt{k}$D. The worst situation for S/I occurs when mobile terminal is at the boundary and placed at one of the corner of the hexagon.

$$S/I = \frac{P/R^4}{\sum_k P/D_k^4} = \frac{D^4}{R^4}\frac{1}{6(1+2/9+3/16\ldots)} = \frac{1}{8.457}q^4$$

Considering only first tier neighbour cells = 6, $D/R > (6 \times S/I)^{1/4} > (6 \times 79.43)^{1/4} \approx 4.67$ and as $q = D/R = \sqrt{3N}$

$$\Rightarrow N = (1/3)(D/R)^2 \approx 7.26$$

Let N = 9, the nearest cluster size, then $q = \sqrt{(3 \times 9)} = 5.196 > 4.95$, with this value $S/I = q^4/8.457 = 86.19$ = 19.35 dB which is greater than the required S/I 19 dB. The quantity η = channel /cell = C/N = 110/9 = 12.22=12 channels/cell. This is to mention worth that single tier approximation is helpful when propagation exponent is less than 4.

Example 3.3 Now consider the frequency reuse factor N =7, for a system that requires S/I 18 dB. Calculate the worst case of signal to interference ratio. Also justify whether N=7 is an acceptable limit for co-channel interference.

Solution For N=7, $q = \sqrt{3N} = 4.6$, if the path loss component κ =4, then from the relation

$q = (N_I \times S/I)^{1/\kappa}$, $S/I = (4.6)^4/6 = 74.62(18.72 \text{ dB})$,

So, S/I is greater than the acceptable limit 18 dB, so N=7 is the justified reuse factor. Now suppose the required acceptable limit is considered as 19 dB, in that case the suitable choice for N would be 9, for which S/I ratio is 19.3 dB as obtained in Example 3.2.

It is to be noted that increasing frequency reuse factor 7 to 9 means reduction of system capacity, channel/cell from 110/7 to 110/9. This is a designer trade off point. So, co-channel interference is a very important parameter that controls link performance and determines cell planning with appropriate frequency reuse factor for providing system capacity.

Adjacent Channel Interference Two independent systems, simultaneously working on the same frequency can operate efficiently if the separation distance that allows sufficient attenuation of their signal strengths. If the distance between these systems gets smaller than the critical limit, the quality of the received signal in both the systems significantly decreases and end-users experience poor signal quality.
In order to describe the quality of a received signal, several different ratios are introduced:

1. S/N - signal-to-noise
2. S/I - signal-to-interference
3. C/I - carrier-to-interference and
4. SI/NAD - signal-to noise and distortion ratio

The selection of a ratio for a given system depends on the system design and the desired results. Furthermore, the interference in radio systems is divided into co-channel interference and non-co-channel interference. Non-co-channel interference includes adjacent channel interference and inter-modulation interference produced by nonlinear elements (e.g., amplifiers).

Adjacent channel interference (ACI) results from signals, which are adjacent in frequency to the desired signal. ACI is caused by imperfect receiver filtering of the adjacent channels signal and leak into the pass band. ACI can be reduced by careful designing of band pass filter at the receiver end, by using proper modulation schemes that have low out of band radiation, by maintaining proper channel interleaving by assigning adjacent channels to different cells and by the use of multiplexing schemes in time and frequency to separate the uplink and downlink.

The problem of ACI is severe if the adjacent channel user is transmitting very close to the subscriber receiver attempting to receive a base station signal on the desired channel. This is known as the **near-far effect**. It occurs when a mobile user close to the base station transmits on the channel close to it and having weak signal. Base station cannot distinguish this desired mobile user from the weak and close adjacent channel mobile.

Suitable channel allocation schemes may also minimize the ACI problem. A cell should not be assigned channels, which are adjacent in frequency, rather keeping frequency separation as large as possible. If the frequency reuse factor is high, i.e., small N, then adjacent channel separation may not be within the acceptable limit. Tight base station filtering is needed when near and far users share the same cell. Usually, base station system uses a high Q cavity to reject adjacent channel interference.

3.1.5 Enhancement of System Capacity: Cell Division

As the capacity of cellular system can be increased by the use of frequency reuse, there are other ways to increase the capacity also as the demand for traffic grows within a given cell. The possible ways for handling increase traffic are addition of new cells, cell splitting and sectoring of cells.

Cell Splitting The cell within a congested area is subdivided into smaller cells (with reduction of cell radius). It requires a new base station and corresponding antenna with lower height and lower transmitting power. This method is basically addition of new cells to meet the additional need for coverage. Though this

method is very effective, it is costly due to the requirements of installation of towers and antennas with their respective cost though flexible way to expand the system.

The subdivision of large cells into small cells implies that the frequency reuse distance becomes smaller and the number of channel within the same geographic area is increased providing large system capacity. But smaller cells also require hand-off for moving users that necessitated large amount of signaling information to be used and the overall spectral efficiency will be reduced.

In the context of cell splitting, the smallest cell is known as pico-cell, the medium size cell is the micro-cell and the original large cell without splitting is called the macro-cell.

As shown in Fig. 3.11, when the center of city is congested with heavy traffic, the cells can be reduced to its 1/4 size. In order to cover the entire service area, approximately four times of smaller cells are needed. The increased number of cells would increase number of clusters. This also reduces blocking probability as the capacity of the cellular system increases the numbers of times those channels are reused. If the cell radius decreases to $R/2$ then the number of small cells within the same area of coverage will be $[R/(R/2)]^2$ x the number of original large cells with radius R.

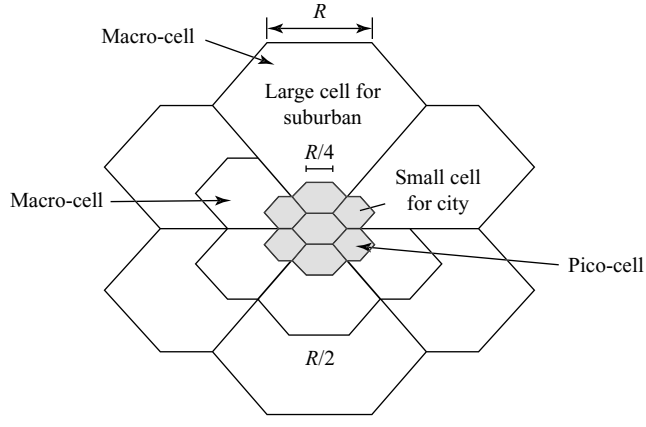


Fig. 3.11 *Cell splitting*

It is to be mentioned that the new smaller micro-cells need low power transmitting base station in order to keep the same received signal strength at the mobile situated at the cell of old and new boundaries. If the path loss exponent κ=4, then as the received power is proportional to $P_{t1}\, R_{\text{cell}}^{-\kappa}$, ($P_{\text{t1}}$ is the base transmitting power), for the new micro-cell structure with cell radius $R_{\text{cell}}/2$, the received signal power would be proportional to $P_{t2}(R_{\text{cell}}/2)^{-\kappa}$.

The ratio of two equal received signal is $(P_{t1}/P_{t2}) = 2^{\kappa} = 2^4 = 16.$

Thus the base transmitting power for smaller cell should be 1/16 of the old cell structure with radius R_{cell}. This is necessary to ensure the same frequency reuse plan as per with the old system.

Cell Sectoring Another way of capacity increase is the sectoring of cells. In general, within a cell the antennas are omni-directional. The set of omni-directional antennas in a cell is replaced by directional antennas with 600 or 1200 aperture. The cell is divided into sectors. This method is more economical, because it uses the existing system structures. This also reduces co-channel interference as with directional antennas, a given cell will receive interference and transmit only a fraction of available co-channel cells. The reduction of co-channel interference is dependent on the angle of sectors. Cell sectoring also splits the channel sets into smaller groups, reducing trunk efficiency.

Consider Fig. 3.12 with 120^0 sectorization. For frequency reuse factor N = 7, there are 6 co-channel neighbors with omni-directional antenna. But for 120^0 directional antenna this co-channel interference reduces to 6/3=2. This also increases signal to interference ratio that can be shown as follows:

We know that, $q = (N_I \times S/I)^{1/\kappa}$, N_I depends on the type of antenna used. For omni-directional antenna with first tier co-channel interferer, N_I =6, but for 120^0 directional antenna it is 2.

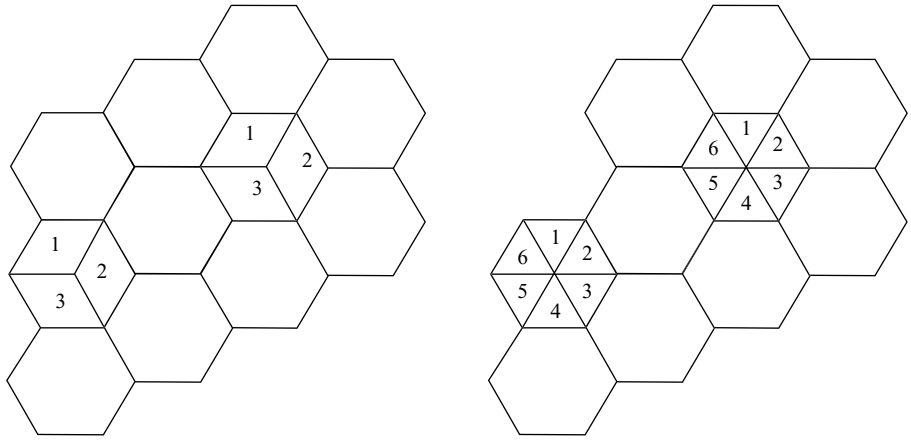


Fig. 3.12 *Sectorization of cells using directional antenna: (a) 3 sectors of 120° each, (b) 6 sectors of 60^0 each*

So, the increase in S/I ratio is then,

$$\frac{(N_I \times S/I)_{120^\circ}}{(N_I \times S/I)_{omni}} = \frac{q^{\kappa}{}_{120^\circ}}{q^{\kappa} omni}$$

$$\frac{(S/I)_{120^\circ}}{(S/I)_{omni}} = 3$$

Thus, S/I ratio increases with number of sectors used instead of omni-directional antenna, but at the cost of extra hand-off that may be needed from sector to sector movement of the mobile user.

Example 3.4 Discuss the effects of the path loss exponent on the frequency reuse for a Cellular system with total of 550 duplex voice channels without frequency reuse. The service area is divided into 152 cells. The required signal-to-co-channel interference ratio is 17 dB. Consider the path loss exponent κ equal to 3,4, and 5, respectively. Also determine

(a) the cell cluster size,
(b) the number of cell clusters in the service area, and
(c) the maximum number of users in service at any instant.

Will there be any effect of base transmitting power with the varied κ when the cell size remains fixed?

Solution We know that

Frequency reuse ratio $q = \sqrt{3N} = (N_I \times S/I)^{1/\kappa} = (6 \times S/I)^{1/\kappa}$

Considering only the first tier co-channel interference cells ($N_I = 6$)

Case 1

(a) For $\kappa = 3$,

$$S/I = 17\ dB \Rightarrow 10^{1.7} \approx 50.12$$

Again,

$$S/I = 1/6 \times q^{\kappa} = (1/6) \times (\sqrt{3N})^3 = (1/6)(3N)^{3/2}$$

$$\Rightarrow 50.12 = (1/6)(3N)^{3/2}$$

or $$N = 14.96$$

The nearest possible values of N will be 16, i.e., the cluster cell size.

(b) Number of cell cluster in service area $= 152/16 = 9.5 \cong 10$.

(c) The maximum number of users in service at any instant = 550 x 10 = 5500.

Case 2: $\kappa = 4$

(a) N = 5.78, cluster size N = 7

(b) Number of cell cluster in service area = 152/7 = 21.714

(c) Maximum number of users in service = 550 x 21.714 =11942

Case 3: $\kappa = 5$

(a) N = 3.26 , Cell cluster size N = 4

(b) Number of cell cluster in service area = 152/4 = 38

(c) Maximum number of users in service = 550 x 38 =20900

Discussion: As the path loss exponent increases, the frequency reuse factor N decreases, this means increase of capacity due to decrease in cluster size. So, the number of users in service also increases. Now regarding the power transmission for fixed cell size (R is fixed), we can proceed in the following way,

We know, $q = D/R = \sqrt{3N}$

As N decreases from 16 to 4 with the increase of κ value from 3 to 5, q is decreased, so for fixed R, D has to be decreased. Again as $S/I = (D/R)^{\kappa}/N_I$, as κ varies from 3 to 5, received S/I decreases in proportion to $\sqrt{3N}$. Considering $N_I = 6$, to keep S/I to the tolerable limit of 17 dB, the transmit power should be increased for fixed R.

Example 3.5 Cell sectoring is one of the important ways for capacity enhancement in cellular system. It also helps in reducing interference. Now, consider a cellular system with 7- cell frequency reuse and a total of 392 traffic channels. Suppose the probability of call blocking is not more than 1%. Assume that every subscriber makes 1 call/hour and on average each call lasts 2.5 minutes, then

(a) Using omni-directional antenna, determine the traffic load in Erlangs per cell and the number of calls per cell per hour.

(b) Repeat (a) for a 120^0 sectoring.

(c) Repeat (a) for a 60^0 sectoring.

(d) Discuss the effect of no sectoring, 120^0 sectoring and 60^0 sectoring on the S/I ratio and on the trunk efficiency.

(e) Determine the minimum frequency reuse factors for no sectoring, 120^0 sectoring and 60^0 sectoring, respectively, taking into consideration that a S/I value of 19 dB or better is satisfactory.

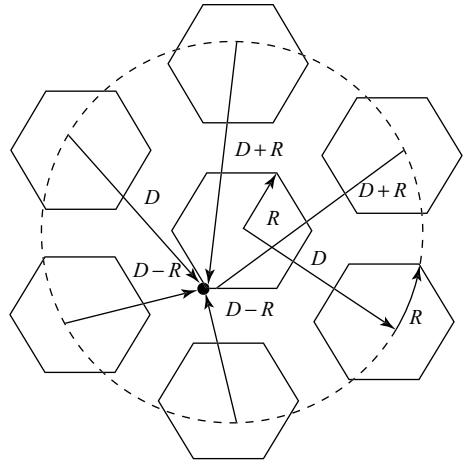


Fig. 3.13 *Worst case co-channel interference scenario*

Solution To solve the problem completely, let us consider the picture for worst-case co-channel interference scenario for single tier cells with N=7.
The number of channels per cell = 392/7 = 56

(a) Here, the blocking probability = 0.01
Consulting Erlang B Table 4 [12] in Appendix C
The number of calls per cell = Traffic load = 43.315 Erlangs
The number of calls per cell per unit time
= Number of calls per cell/Average call holding time
= 43.315/2.5 = 17.326 calls/cell/min
So, calls/hour = 17.326 x60 = 1039.56

(b) For 120^0 sectoring, number of channels/sector = 392/(7x 3) = 18.667≈18
For blocking probability of 0.01 and 18 channels, using Erlang B formula in Appendix C
Number of calls per sector = 10.437
Number of calls per cell = 10.437 × 3 = 32.31 = Traffic load
Number of calls per cell per unit time = Number of calls per cell/Average call holding time
= 32.3/2.5 =12.92 call per cell per min
= 775.2 calls per cell per hour

(c) For 60^0 sectoring,
Number of channels per sector = 392/(7 × 6) = 9.333=9
For blocking probability of 0.01 with 9 channels from Erlang –B Table, we have
Number of calls per sector = 3.78
Number of calls per cell = 3.78 × 6 = 22.68 = Traffic load
Number of calls per cell per unit time = Number of calls per cell/Average call holding time
= 22.69/2.5 = 9.072 calls per cell per min
= 544.32 calls per cell per hour
For omni-directional antenna, $N = 7$, $q = 4.58$

$$\text{S/I} = (1/6) \times q^4 = (1/6) \times (\sqrt{3N})^4 = 73.56$$

$$(S/I)\text{dB} = 18.66 \text{ dB}$$

(d) For worst case co-channel interference when the mobile is at the cell boundary (as shown in Fig 3.12) with $\kappa = 4$, we can use the accurate formula for S/I as,

$$(S/I)=\frac{R^{-\kappa}}{2(D-R)^{-\kappa}+2D^{-\kappa}+2(D+R)^{-\kappa}}=\frac{1}{2(q-1)^{-\kappa}+2q^{-\kappa}+2(q+1)^{-\kappa}}=53.28$$

$(S/I)_{dB} = 17.26$ dB (this is valid of omni-directional cell)

Now for 120° sectoring, the number of interferer cells reduces from 6 to 2.

$$(S/I) = 1/2\, q^4 = 1/2\times(\sqrt{3N})^4 = 220.50 = 23.43 \text{ dB}$$

Using accurate formula for the S/I ratio for $120° = \frac{1}{q^{-4}+(q+0.7)^{-4}} = 24.6$ dB

and for 60^0 sectoring, $(S/I)_{60^0} = q^4 = 26.4$ dB

Again, trunk efficiency = carrier traffic intensity in Erlangs per channel
= Traffic load/ Number of channel per cell

Case 1: Traffic efficiency for omni-directional antenna = 43.315/56 = 0.77
Case 2: Traffic efficiency for 120^0 sectoring = 32.31/56 = 0.55 and
Case 3: Traffic efficiency for 60^0 sectoring = 22.68/56 = 0.405

From the results, it is very much clear that sectoring of cells reduces interference and hence signal-to-interference ratio increases as the number of sectors increases compared to omni-directional antenna without sectoring. Again, if the total number of channels available for each cells to be partitioned into sectors, the trunking efficiency for each cell is reduced than that of without sectoring

(e) **Case 1: For no sectoring**

With $(S/I) = 19$ dB $\Rightarrow (S/I) = 79.4$

Again,

$$(S/I)= (1/6)\, q^k = (1/6)\times(\sqrt{3N})^4 = 79.4$$

$$(3N)^2 = 79.4\times 6 = 476.4$$

Or $$\Rightarrow N = 7.27 \approx 7$$

Let N = 9 be the minimum frequency reuse factor, then $(S/I) = 20.84$ dB which is greater than 19 dB and will be satisfactory.

So, N = 9 is the minimum frequency reuse factor.

Case 2: For 120° sectoring

$$S/I = 1/2\, q^4 = 1/2\,(3N)^2 = 79.4$$

$$\text{Or, } (3N)^2 = 79.4\times 2 = 158.8$$

$$\Rightarrow N = 4.2 \approx 4$$

Case 3: For 60° sectoring

$$S/I = q^4 = (3N)^2 = 79.4$$

$$\Rightarrow N = 2.9 \approx 3$$

So the minimum frequency reuse factor for no sectoring, 120° sectoring and 60° sectoring are 9, 4 and 3 respectively.

Example 3.6 If 20 MHz of the total bandwidth is allocated for a duplex channels in a cellular wireless system with frequency reuse factor 4 and each simplex channel is 25 kHz radio frequency bandwidth, and then find

(a) the available number of duplex channels,
(b) the number of channels per cell, and
(c) if the system allocates 40 kHz bandwidth for guard band, then what will be the effective number of available duplex channels for voice communication?

Solution

(a) Number of duplex channels = 20 MHz / 2 x 25 kHz = 400 channels.
(b) Total number of channels per cell = 400/N = 400/4 = 100.
(c) Number of duplex channels considering 40 kHz guard band = $(20000 - 40) / 50 \cong 399$.

Example 3.7 The first generation cellular AMPS system is allocated 25 MHz band with 21 control channels for one direction communication at 800 MHz range. The total bandwidth required for communication is 50 MHz. The frequency separation between the uplink and downlink is 45 MHz. If there is 40kHz guard bands then how many total numbers of channels within the system if each simplex channel occupies 30 kHz? Now if the base station transmits control information on channel 352 at 880.560 MHz, then determine the following:

(a) Whether the AMPS system is simplex, half-duplex or duplex.
(b) What is the transmission frequency of a mobile subscriber unit transmitting on channel 352?
(c) Determine the number of voice channels and control channels for each carrier if the A-side and B-side cellular carriers are evenly distributed.
(d) If frequency reuse factor N=7 for hexagonal cell structure having equal coverage area, find the distance between the two nearest co-channel cells. Repeat it for N=4.
(e) Considering path loss exponent $\kappa = 4$ and first tier interfering cells only, find the minimum signal to interference ratio S/ I to be maintained for the system with N =7 and N =4.
(f) If the AMPS system is defined with acceptable limit for S/I ratio as 15 dB, then what will be the suitable frequency reuse factor N?

Solution

(a) Given the total bandwidth B = 50 MHz and one-way communication requires 25 MHz bandwidth so that system is full duplex. The AMPS system is duplex because we know each channel of AMPS is 30 kHz. The 60 kHz bandwidth is split into two simplex channels for reverse channel (Mobile-to-base station) and forward channel (base station to mobile) each with bandwidth 30 kHz. The forward channel is separated by 45 MHz frequency bandwidth with the reverse channel.
Bandwidth of duplex channel is =2 x 30 kHz = 60 kHz
Numbers of channels N_{ch} = (50 MHz - 40 kHz) / 60 kHz = 832
(b) We need to calculate the reverse channel frequency for the subscriber.
$f_{for\text{-}ch} = 880.560$ MHz
$f_{rev\text{-}ch} = 880.560 - 45 = 835.560$ MHz
(c) Given N= 832, and total number of control channel = $N_{control}$ = 2 x 21 = 42
So, the number of voice channels for each carrier = N_{voice} /2 = 790/2 =395 channels
Number of control channels for each carrier = $N_{control}$/2 = 42/2 = 21 channels

(d) For N = 7 , q = D/R=$\sqrt{3N} = \sqrt{21} = 4.58$

So, the co-channel distance D = 4.58 R, R is the cell radius

If N = 4, $D = \sqrt{3N}\ R = 3.46\ R$

(e) With N=7, q =4.58 = $(N_I \times S/I)^{1/\kappa}$, N_I = number of first tier neighbor cells = 6

$$\Rightarrow (S/I) = (4.58)^4 / 6 = 73.334 = 18.65 \text{ dB}$$

With N =4, $S/I = (3.46)^4/ 6 = 23.88 = 13.78$ dB

$$(S/I) = (3.46)^4 / 6 = 23.88 = 13.78 \text{ dB}$$

So, the minimum S/I ratio for N = 7 would require greater than 18.65 dB and for N=4 it would be greater than 13.78 dB.

(f) Because for N= 4, the S/I ratio is 13.78 dB which is less than the acceptable limit of 15 dB, N = 4 is not the suitable solution. But for N=7, the S/I ratio is 18.65 dB which is above the acceptable limit, so N =7 would be the suitable choice.

Example 3.8 Consider a metropolitan city with total area of 2500 km^2 is covered by hexagonal cellular cell array with 7 cells reuse pattern. The radius of each cell is 6 km. The cellular system has been allocated total bandwidth of 25 MHz and full duplex channel bandwidth of 30 kHz, total 40 kHz guard band is used as in FDMA system, there are total 16 control channels. Determine

(a) the number of cells in the service area,
(b) the number of channels without frequency reuse, and
(c) the cell capacity.

Solution Consider Fig. 3.14 of a hexagonal cell of radius 6 km.

The length AP = AO sin 30° = 3.0 km, where AO = 6 km radius of cell.

OP = AP cos 30° =3 $\sqrt{3}$ km

Area of Δ AOB = $1/2 \times AB \times OP = 9\sqrt{3}\ km^2$

Area of the hexagon

$= 6 \times \Delta AOB = 6 \times 9\sqrt{3} \cong 93.53\ km^2$

Total service area $= 2500\ km^2$.

Number of cells within the service area = 2500/93.53 = 26.72 ≅ 26

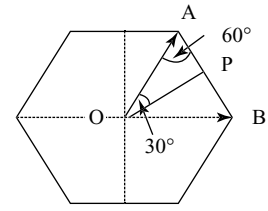


Fig. 3.14 *Hexagonal cell with 6 km radius*

(b) Number of channels without frequency reuse = (Total bandwidth – Guard bandwidth) / channel bandwidth = (25000 – 40) / 30 = 832 channels

(c) Number of data channel per cluster = total channels – control channels = 832 –16 = 816

Cell capacity = Number of available channels per cell = 816/7 = 116.57

3.2 CHANNEL ASSIGNMENT SCHEMES IN CELLULAR NETWORKS

The Channel Assignment (CA) scheme determines an assignment of channel(s) to base stations such that frequency reuse is maximized for a given set of channels or frequencies (F) and a set of base stations (B) in the coverage area.

Several channel assignment schemes are there as given in Fig. 3.15.

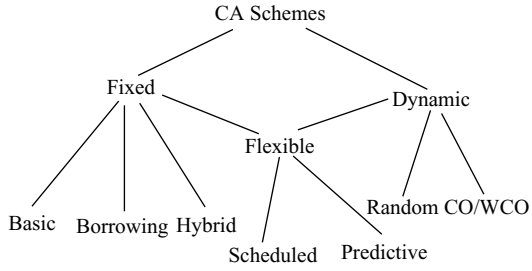


Fig. 3.15 *Different channel assignment schemes*

3.2.1 Fixed Channel Assignment (FCA)

In this scheme, a fixed set of channels is permanently assigned to each cell and they are reused in the co-channel cells. A user is assigned an unoccupied channel on demand, which is relinquished after the call is over. If the number of calls exceeds the channel set for a cell, the excess calls are blocked. Various graph coloring techniques can be used to maximize reuse of the available frequency channels. Static channel assignment schemes perform well under heavy traffic conditions.

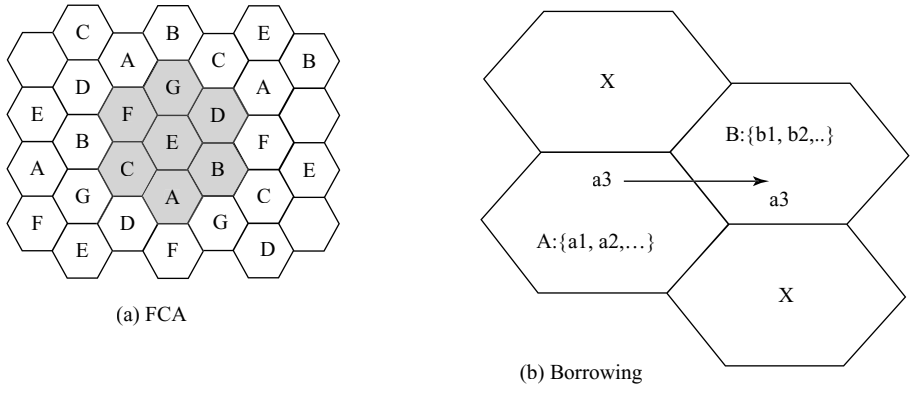


Fig. 3.16 *Fixed channel assignment schemes*

In FCA, a channel can be borrowed from neighboring cell if no interference with existing calls happened. Borrowed channel is locked in non-co-channel cells of borrower cell. But in this simple borrowing scheme, channel-blocking performance suffers under heavy traffic conditions. In Fig 3.16(a), channels are permanently assigned to A-G cells, whereas in Fig. 3.16(b), channel a3 is borrowed from cell A to cell B.

Flexible Borrowing Some fixed channel set of a cell is divided into two groups: One group for local uses only, other group of channels for borrowing. The number of channels in each group is determined a priori depending on traffic conditions.

Borrowing with Channel Ordering It is the extension of flexible borrowing scheme, where the number of channels in both groups can vary dynamically depending on the traffic condition. Each channel is ordered, the first channel has highest priority of being locally used and the last channel has highest priority of being borrowed. Ordering depends on traffic pattern within the cells. The higher order channel is released first and reallocated to ongoing call using lower order channel to unlock the borrowed channels.

3.2.2 Dynamic Channel Assignment (DCA)

In DCA, no cell has a proprietary set of channels, instead channels are allocated to users on demand from a central pool based on a cost function. DCA can again be divided into three categories:

1. **Random DCA:** Here available channels are randomly assigned. This scheme has poor channel utilization.
2. **Channel Ordering (CO):** In this a cell can use any channel, but each has a different ordering. Channel with the highest priority is selected for the cell.
3. **Weighted Carrier Ordering (WCO):** Here each cell develops "favorite" channels from past experience. This scheme adapts faster to traffic changes than DCA and CO, but needs more time to search for the highest priority channel that causes delay.

3.2.3 Hybrid Assignment (HA)

In HA, a set of permanent (fixed) channels is allocated to each cell. For channel shortage in a cell, channels are assigned from a set of flexible channels according to a DCA strategy. Flexible channels can be distributed among cells in a scheduled or predictive manner.

Scheduled distribution assumes knowledge of future changes in traffic distribution, while predictive scheme continuously monitors the traffic in each cell so that flexible channel reallocation can be done at any time.

Comparison of CA Schemes

Fixed Channel Assignment (FCA)

1. Not flexible, poor utilization
2. High blocking rate for non-uniform traffic
3. Low computational cost
4. Requires frequency planning
5. Not desirable in micro-cell architectures

Dynamic Channel Assignment (DCA)

1. High utilization of channels
2. High computational complexity
3. Suitable for non-uniform traffic and micro-cell architectures

Hybrid Channel Assignment (HCA)

1. Combination of FCA and DCA
2. Converges to FCA performance under heavy traffic

3.3 CELLULAR COMMUNICATION PRINCIPLE

3.3.1 Network Architecture

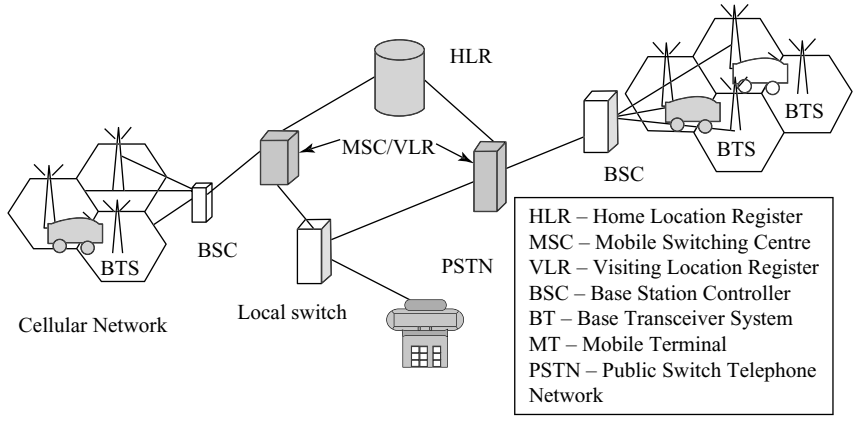


Fig. 3.17 *Cellular communication network architecture*

The identity of the mobile users and their service information are kept in a database of a **home location register known as HLR**. The regional register in the visited domain is known as VLR. Users subscribe to a regional sub-network for communication services, called **home network (HN)**. The other network premises within which users roam from its HN is called visited network or foreign network. VLR is the foreign network database register that updates the location information of the user with its HLR. **MSC** controls BS and serves as gateway to the backbone network (PSTN, ISDN, and Internet).

The association between the MN (Mobile Node) and its home network is made through **registration update process**. When mobile moves from home network, it still maintains association with it through registration process. Before the registration update, HLR checks the **authenticity** of the mobile. This is necessary for security measure. Home network must know the current location of the mobile at all time for **message delivery**. MSC takes part for registration update, authentication and call delivery process. Authentication typically depends on exchanges of cryptographic messages between the communicating parties. A key is needed to open the lock in order to read the message like password. The mobile's identity is its permanent address, only the mobile and its HLR know the mobile's permanent address . Similarly, the identity is locked in a box; only the mobile and HLR know the key to open it.

A base station employed as a hub to handle the information transfer between a source and a destination user via. the mobile terminal (MT). Due to the limited coverage of a cell, wireless network needs a wire line (PSTN) or a satellite network for extended coverage globally. The network with the wide area coverage is called a Personal Communication Network (PCN). In general PCN comprises both wire line and wireless links interconnecting traffic handling devices refereed to as routers.

3.3.2 Mobility Management

Most salient feature for wireless communication is the flexibility to support user roaming. Though the frequency reuse enhances system capacity, it generates the need for **Mobility Management (MM)** while MT crosses the cell boundary. MM is a two step processes: Hand-off Management and Location Management both of which can be defined as follows:

1. **In-session mobility management:** Move during an active call that necessitates handoff management.
2. **Out-of-session mobility management:** Move in standby mode, necessitates location management.

To facilitate user roaming, there must be effective and efficient handoff mechanism so that the mobile's connection with its current serving base station is handed off to its target base station to maintain service continuity.

Hand-off Process When mobile users move into a different cell while a conversation is in progress, the MSC automatically transfers the call to a new channel belonging to the new base station. This hand-off operation not only involves identifying new cells, but also requires that the voice and control signals be allocated to channels associated with the new base station. Hand-off process consists of two phases such as **initiation phase** and **execution phase.**

A handoff is initiated when average signal level from the new BS exceeds that from the current BS by a specific threshold. Normal acceptable voice quality signal level lies between –90 dBm to –100 dBm. A slightly stronger signal level is used as a threshold for handoff. The execution phase will include the allocation of new radio resources channel and exchange of control messages. Figure 3.18 describes the hand off process when mobile moves from cell A to cell B.

Handoff must be performed successfully and as frequently as possible. System designer must specify an optimum signal level at which to initiate a handoff. In general, the current BS monitors the RSS quality from mobile station, when RSS goes below a specified threshold, BS instructs the mobile to measure signal strength from neighboring BS's. Two situations may arise:

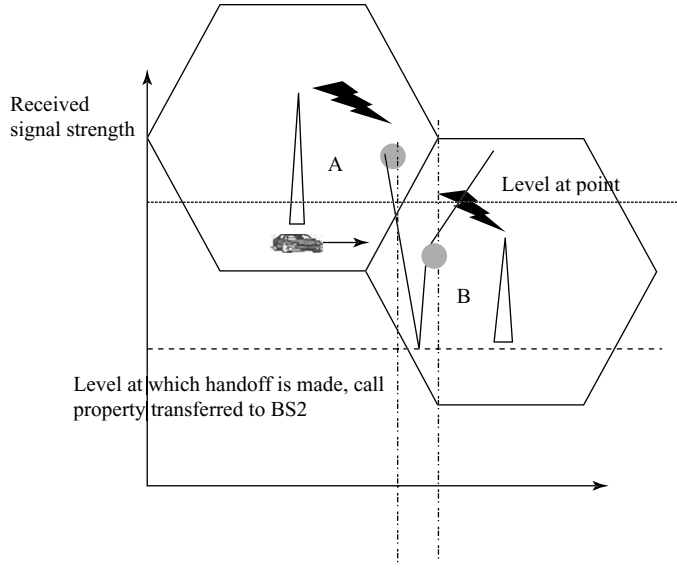


Fig. 3.18 *Hand-off process due to user movement*

Case 1: Mobile station sends collected information to BS

1. BS conveys the signal information to its MSC, which selects the most suitable BS for the mobile station.
2. The selected BS and the mobile station are informed when the new BS assigns an unoccupied channel to mobile station.

Case 2: Mobile station itself selects the most suitable base station

1. Mobile station informs the current BS, who conveys information about the new BS to its MSC
2. Selected BS is informed by MSC, which assigns a new channel.

In the first generation analog cellular system, signal strength was measured by the BS and supervised by the MSC. Each BS continuously monitors the signal strength of all received signals from the mobile stations and then estimates the location of the mobiles from the BS.

Whereas in the second generation digital TDMA cellular network like GSM, handoff decision is mobile assisted called MAHO. Each mobile station continuously measures the received power from the neighboring BSs and then sends report to the serving BS. Handoff is initiated when the received signal power from neighbor BS exceeds the power of the serving BS. Obviously, MAHO is faster than the first generation handoff.

Two basic hand-off types are there. These are:

1. **Hard handoff:** A mobile having a radio link with only one BS at any time characterizes this type of handoff. The old connection is terminated before a new connection is activated. This mode of operation is referred to as **break before make.** Second generation mobile communication system, based on GSM, falls in this category. Hard handoffs are of two types: (i) Intra-cell handoff (i.e., mobile movement inside the cell) due to deteriorated channel quality or resource rearrangement. (ii) Inter-cell handoff caused by mobile movement away from cell.
2. **Soft Handoff:** The mobile can simultaneously communicate with more than one BS during the handoff. New connection is made before breaking the old connection and is referred as **make before break**. Second and third generation mobile communication systems, based on CDMA use soft handoff.

Handoff can be categorized in many ways. These are 1. Transparent handoff (TH), 2. Subscriber assisted handoff (SAH), 3. Inter-system handoff, 4. Mobile assisted handoff (MAHO).

1. **In transparent handoff:** The user will not be aware of the handoff process.
2. **Subscriber assisted handoff:** It Helps the network to inform a user in advance that the call will be dropped if the movement continues so that the user can determine what to do accordingly.
3. **Inter-system handoff:** It occurs if mobile move from one cellular system to another controlled by a different MSC during the course of a call.
4. **Mobile assisted handoff:** It is initiated when the power received from the BS of a neighboring cell begins to exceed the power received from the current BS for a certain period of time. In a mobile-assisted hand-off process, the MS makes measurements and the network makes the decision. In the circuit-switched GSM network, the BSC is in charge of the radio interface management, means allocation and release of radio channels and hand-off management. The hand-off time between hand-off decision and execution in such a circuit-switched GSM is approximately one second.

It is also possible to calls by using very stringent conditions on new calls. As for example, higher threshold value is considered for the new calls. For giving priority to handoffs, use of **guard channel** is very common where a fraction of the total available channels in a cell is reserved exclusively for hand-off requests from ongoing calls, which may be handed off into the cell. However, reserving guard channels reduces spectrum efficiency for fixed channel assignment scheme. Use of guard channels with dynamic channel assignment may lead to efficient spectrum utilization by minimizing number of required guard channels according to the demand. **Queuing of handoff calls** is another method to decrease the forced termination of a call and is used to control call-dropping probability and total carried traffic. Queuing of handoffs is possible due to the fact that there is a finite time interval between the time the received signal level drops below the handoff threshold and the time the call is terminated due to insufficient signal level. The delay time and size of the queue is determined from the traffic pattern of the service area.

3.3.3 Location Management

The problem of Location Management (LM) involves the tracking of a mobile station to route incoming call requests within an allowable time constraint.

While enjoying the freedom of being mobile, the user creates an uncertainty about the exact location of the mobile station. Unless controlled, uncertainty may grow without bound. Here lies the importance of the LM.

LM has two parts:

1. Registration Update or Location Update (LU) which is device centric, and
2. Paging which is network centric.

LU is a process to generate the location updates by informing the mobiles current location whenever it moves into new **Location Area (LA).** Through this LU process mobile tells the network "I am here". By monitoring the beacon signals from the BS, mobile senses new LA and sends registration update request to BS of the current cell that contains mobile identity number. BS forwards this registration request to MSC, which in turn update VLR, VLR adds the records of MN and forwards the request to HLR. HLR performs authentication, and stored the new VLR information, sends acknowledgement to new VLR. After that registration cancellation is sent to old VLR, old VLR removes the records and sends cancellation acknowledgement to HLR.

Paging is a means of locating a mobile within a LA. When new session arrives, network system searches the mobile with "Where are you?" A commonly used approach is polling; the called MSC broadcasts a polling message which contain **MIN (Mobile Identity Number)** of the called mobile. The concern BS of a cell relays the paging message to the mobile terminals, the called mobile responds to its MSC via. the current BS that contains BS ID. Paging systems vary widely in their complexity and coverage area. Simple paging covers a range of 2 km to 5 km or may even be confined to within a building.

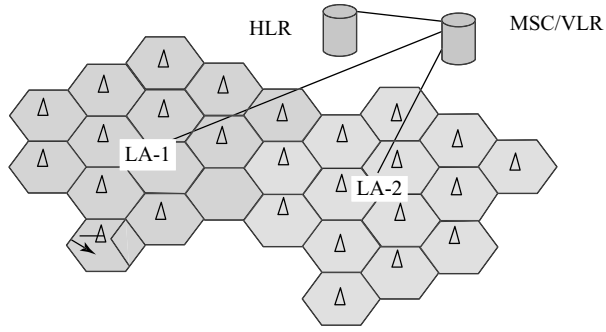


Fig. 3.19 *Location area in a cellular network*

Paging cost is proportional to number of calls arrived and number of cells paged for each call, whereas the LU cost is proportional to frequency of location update. Too many LU incurs high update cost and low paging cost. Too few update cost increases paging cost and decreases update cost. So tradeoff between the LU cost and paging cost is required.

Types of Mobility

1. **Terminal (Device) mobility:** Terminal mobility enables the network to route calls to mobile terminals regardless of the point of attachment.
2. **Personal mobility:** It is the ability to use the same terminal by different users at the same time through different addresses using Universal Personal Telephone Number (UPT).
3. **Service/Session mobility:** Session mobility allows users to roam beyond own networks. It is the ability to move the complete set of services from one network to another.

3.4 RADIO RESOURCE MANAGEMENT

Managing Radio Resource (RR) plays a great role in cellular communication to maintain certain Quality of Service (QoS). RR Management (RRM) encompasses frequency and or time channels, power transmission and access to base stations in order to control some amount of available radio resources. At the same time, it also takes care of users QoS satisfaction. From the network point of view, some objectives, such as total network throughput, total resource utilization and earning revenue are set. The performance of individual user depends on the good RRM techniques. CAC is one of the ways to support RRM where the decision to grant or reject an arriving call is absolutely based on network loading condition. Thus, traffic of the admitted calls is controlled by RRM techniques. This includes scheduling, handoff and rate of power control schemes. The problem of CAC in wireless networks is more than wire line due to channel impairment, channel access techniques, interference from other mobiles or neighbor BS and over all the limited radio resource bandwidth. In the wireless networks frequency planning, access method for air interface, etc. play an important role to RRM. Within a geographical area, a network planner wants to cover the entire area with minimum number of BS to make it cost effective. Again within a single BS, the desirable objective is to maximize the access of large number of mobile users. Here, lies the importance of suitable CAC scheme that may help RRM techniques to utilize available radio resources in a beneficial manner.

3.4.1 Call Admission Control

The Call Admission Control (CAC) protocol is one of the most important engineering issues in cellular as well as wireless networks since it belongs to the category of resource management. CAC protocol means control of the admission of calls into a particular cell to provide the desired Quality of Service (QoS). It is based on bandwidth adaptation, call admission test and next cell prediction of the call. CAC protocol should efficiently support handoff and also maintain high resource utilization of the network.

The hand-off dropping is much annoying for the users than a new call being blocked. In order to prevent this situation, the system usually reserves some of its resources in order to give privilege to the handoff users. On the other hand, if the system reserves resources too much, it is possible that in case the system resources not are being used by any user still can get refused call request. This kind of waste is also undesirable from service provider point of view. So, it is the requirement to develop good CAC protocols to determine the reasonable usage of the system resource and keep good balance. To provide the desired QoS in cellular networks and as well as wireless networks, CAC protocol plays a significant role. Mobile communication researchers all over the world are experimenting with several CAC protocols to reduce congestion and to get better QoS.

The role of CAC protocols is to

1. guarantee the signal quality,
2. guarantee the low call dropping probability,
3. giving priority to some classes,
4. maximize revenue,
5. fair resource sharing,
6. guarantee rate of transmission, and
7. guarantee packet-level QoS.

CAC plays an important role in QoS provisioning in terms of signal quality, call blocking, call dropping probabilities, packet delay and loss and also to maintain the transmission rate. For 1G and 2G cellular networks single voice service was the main target. In the evolutionary process of 3G networks, the importance of CAC also changes from voice to multimedia services such as voice, video, data and audio together with varied QoS requirements.

CAC Variants

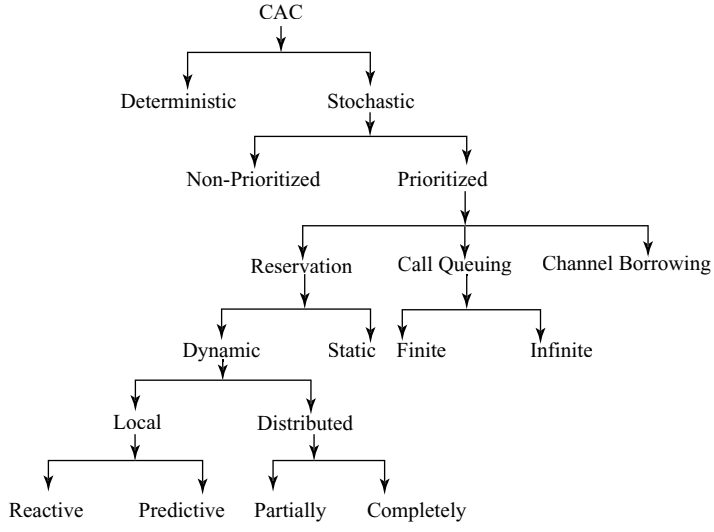


Fig. 3.20 *Different types of CAC protocols*

In general, CAC schemes are mainly of two categories (a) Deterministic CAC, and (b) Stochastic CAC. In deterministic schemes, QoS parameters are guaranteed with 100% confidence, which require knowledge of user mobility, which is not practical. In stochastic schemes, QoS parameters are guaranteed with some probabilistic confidence.

This stochastic scheme achieves higher resource utilization than deterministic schemes. Stochastic schemes are of two types: (i) Non-Prioritized, and (ii) Prioritized. The prioritized scheme is further divided into three types: (a) Channel borrowing scheme (b) Call queuing scheme, and (c) Reservation scheme. Channels are made available at each cell by channel assignment schemes: (i) Fixed Channel Assignment (FCA), (ii) Dynamic Channel Assignment (DCA), and (iii) Hybrid Channel Assignment (HCA) schemes.

In FCA scheme, a set of nominal channels is permanently assigned to each BS. A new call can only be accepted if there is a free channel available in the cell. Several variations are there in FCA, one approach is called the **channel borrowing strategy**. In a channel borrowing scheme, a cell that has used all its assigned channels, can borrow free channels that does not interfere with existing calls. A cell is allowed to borrow channel from a neighboring cell, if all of its own channels are already occupied. In DCA, all channels are kept in a central pool to be shared by all calls in every cell. DCA is flexible but it is less efficient in heavy traffic. HCA, combination of FCA and DCA schemes, overcome these drawbacks.

In call queuing scheme, queuing of hand-off requests, when there is no channel available can reduce the dropping probability at the expense of higher new call blocking probability (CBP). The queue can be of finite length which is more realistic than infinite length queue concept. In finite queue scheme, hand-off calls waiting in the queue have priority over new calls waiting in the queue to gain access to available channels.

Reservation schemes are of two types: (i) Static Reservation, and (ii) Dynamic Reservation. CAC in static reservation scheme keeps a fraction of the total available channel reserved permanently in a cell for hand-off request from ongoing calls, which may be handed off into the cell. Here, hand-off CDP will be decreased whereas new CBP will be increased. The static reservation scheme results in poor resource utilization. To overcome this drawback, dynamic reservation scheme is preferred. There are two types of dynamic reservation schemes: (i) Local reservation, and (ii) Distributed reservation. In local reservation scheme, each cell estimates the state of the network using local information only, while in distributed schemes each cell gathers network state information in collaboration with its neighboring cells.

Local admission control schemes are again of two types (i) Reactive, (ii) Predictive. In reactive approaches, admission policies adjust their decision parameters, i.e., threshold and reservation level as a result of an event occurrence, such as call arrival, completion or rejection. Predictive approaches predict future events to prevent undesirable QoS degradations. Guard Channel (GC) scheme falls in reactive approaches and performs remarkably well when load does not differ from the expected level. Local admission control is very simple but they suffer from the lack of global information about the changes in network traffic. Distributed admission control accesses global traffic information with increasing complexity and overhead due to information exchange between cells. To overcome these drawbacks predictive approaches are better. Distributed CAC (DCAC) schemes can be categorized into (i) Partially, (ii) Completely Distributed. In partially distributed approach, all the necessary information are gathered from a set of neighboring cells but the final decision is made locally in the network controller. In completely distributed approach, all the necessary information are gathered from a set of neighboring cells and neighboring cells are involved in the decision making process.

A good DCAC scheme should efficiently support handoffs and maintain a high utilization of the scarce radio resources with guaranteed of QoS in cellular networks. If the handoff occurs frequently, cellular/wireless networks welcome a new paradigm in the area of network congestion and admission control. The reason behind this is that the cellular/wireless mobile users may change their radio cells many times during the lifetime of connection and as a result connection set-up time does not guarantee the resources throughout the lifetime of connection. In DCAC scheme, it is assumed that the new call arrivals admitted to any cell are Poisson whereas the Enhanced DCAC scheme (EDCAC) considers non-Poisson call arrival process and also considers difference on mobility support for low and high mobility calls. Network congestion is difficult to resolve when real-time traffic is present without sacrificing the desired QoS. Mobility of users adds challenging requirements to the CAC protocols.

DCAC scheme reserves an amount of resource for potential handoff when accepting a new call. However, the potential movement of a user can not always be known by the CAC controller. Therefore, it is possible that an active mobile user may move to a radio cell, which has no adequate resource for handoff. In that

situation, QoS will be degraded or the call will be dropped due to the continuation of movement. Sometimes users prefer to control movement to maintain the communication in progress. But the user cannot do with the Transparent Handoff (TH) approach because no information is fed back to the user when a failing handoff is going to be happened. As a result degradation of QoS or call dropping probability increases. DCAC scheme can be considered as TH scheme when a traffic profile factor is considered in DCAC scheme. When a failing handoff occurs due to user's movement, Subscriber Assisted Handoff (SAH) approach is used. In SAH, the subscriber should be informed in advance so that (s)he can decide whether or not to control movement since a subscriber may sacrifice mobility for maintaining communication in progress.

3.5 PRIORITIZATION OF HANDOFF CALLS AND PERFORMANCE ANALYSIS OF CELLULAR COMMUNICATION SYSTEM

With the development of 3G cellular communications, there is a need for efficient and faster handoff with minimum call dropping probability. As the number of mobile user increases with the demand of both voice and multimedia data services, the demand for efficient handoff algorithms also increase. Handoff call queuing is an important approach for better performance in cellular network. Prioritization of handoff call over originating call **using different queue size** is the important technique for evaluating system performance. Higher priority will be given statically or dynamically to the handover calls based on different criteria. Due to the increased traffic in cellular mobile networks, the availability of service depends on the availability of free channels and on the traffic profile within the system. From the traffic point of view, QoS is determined by the two quantities, **probability of call blocking** and **probability of forced termination**. Tradeoff is always required between these performance measures and the system configuration parameters. Giving priority to hand-off calls over new originating call can decrease the probability of forced termination.

Hong and Rappaport [D. Hong 1986] proposed a traffic model for a hexagonal cell. Both the originating and hand-off calls within a cell are generated according to Poisson process, with mean rates λ_O and λ_H respectively. The service area is assumed to be of homogenous topology. A cell within the area is marked and handoff request is generated in the marked cell. When a moving user holding a channel approaches from a neighboring cell with signal strength below the handoff threshold, a handoff request is generated in the marked cell. All the handoff and new calls generated are independent to each other. There are numerous methods for performing handoff depending upon the kind of network entities and several traffic models have been established based on different assumptions about user mobility.

Three phenomena can occur, when calls arrive to the system, they can be either (i) Blocked, because of unavailability of channel (or queue is full in case of call queuing provision) at the time of request, (ii) Dropped, if waiting time exceeds the maximum time-out before getting the service due to energy drops below the minimum, and (iii) Served by the system and completed within the cell if the call holding time is less than the cell residence time, otherwise call is left for the adjacent cell. Obviously, performance of all those events depends on the three traffic characteristics: (a) Call arrival rate, (b) Call holding time, and (c) Cell residence time.

1. **Call arrival rates (λ):** New call request or handoff call request arriving at every cell follows Poisson distribution and all call generation are independent in nature.
2. **Call holding time (t_c):** The call holding time of a call is the time duration between the beginning of a call and the completion of a call. It is assumed that call holding time is exponentially distributed with probability density function $f(t_c) = \mu_D\, e^{-\mu t}$ and mean μ^{-1}.
3. **Cell residence time (t_r):** Cell residence time is the residing time of a mobile node within the cell's coverage area. It is also assumed that cell residence time is exponentially distributed with probability density function $f(t_r) = \mu_D\, e^{-\mu_D t}$ and mean μ_D^{-1}.
4. **Channel holding time (t_h):** If the mobile node given a channel, this channel would be released either by the completion of the call in the cell or by a handover process to a neighboring cell. Thus the channel

holding time is less than the call holding or cell residence time. Channel holding time is defined as the min (t_c, t_r) and is exponentially distributed with a mean $\sigma^{-1} = (\mu_D + \mu)^{-1}$.

5. **Queuing time (t_q):** It is the waiting time (patience time) in the queue before getting served. Time-out period determines the maximum time a call stays in a queue before departing from the system. Hand-off request and new call request are subjected to time-out period and are exponentially distributed with mean γ_h^{-1} and γ_n^{-1} respectively.

3.5.1 Performance Parameters for Wireless Networks

There are some standard parameters those are used for the performance measure of any network, Call Blocking Probability ($\mathbf{P_B}$), the Probability of Forced Termination ($\mathbf{P_F}$), Grade of Service (GoS), Ratio of Carried Traffic to the Total Offered Load, which is defined as Traffic Utilization. A call which enters into the service, has two possible outcomes. One is that the call is completed successfully, the other is that the call is forced to terminate prematurely because the mobile node experience unsuccessful hand-off attempt prior to completion.

Let the probability for call blocking is P_b. It is the probability that a new call attempt does not get service. The blocking probability of a new call is the sum of two terms.

P_B = (Number of new call blocked + Number of New call timed out) / Number of new call arrived

The probability of forced termination $\mathbf{P_f}$: Mobile node may experience unsuccessful handover, so

P_F = (Number of hand-off call blocked + Number of hand-off call timed out) / Number of hand-off call arrived

Grade of Service GoS is a cost function that penalizes the fact that handoff forced termination probability is much more annoying than the new call blocking and it is used as a reference of the GoS offered by the system. GoS is expressed as

$$GoS = P_B + 10\ P_F$$

Offered load to the system is the sum of new calls and hand-off calls that have arrived and is equal to $\lambda = \lambda_n + \lambda_h$

The total carried traffic is the amount of traffic admitted to the cellular network as opposed to the offered load. In general, the carried traffic is less than the offered load because of the new call blocking and handoff forced termination probability. For any system, it is desirable that the percentage of the offered load that is carried is certainly desired to be as high as possible. This percentage would decrease with the increase of offered load and probability of call blocking and forced termination. Thus,

$$\text{Total carried traffic } = [(1 - P_F)\ \lambda_n + (1 - P_B)\ \lambda_n]$$

So, **traffic utilization** (Trunking Efficiency) is defined as the ratio of the total carried traffic to the total offered traffic.

$$\text{Traffic utilization } = [(1 - P_F)\ \lambda_n + (1 - P_B)\ \lambda_n] / (\lambda_n + \lambda_h)$$

The probability of forced termination can be decreased by giving priority to hand-off attempts over new call origination. A subset of channels is exclusively used for hand-off calls. In the first priority scheme, a hand-off call is terminated if no channel is immediately available in the target cell. In the second priority scheme, a hand-off attempt is held in a queue until either a channel becomes available for it, or the received signal power level becomes lower than the receiver threshold level.

Priority of Hand-off Call: Scheme I In this scheme priority is given to hand-off requests by assigning $\mathbf{S_C}$ channels exclusively for hand-off calls among the **S** channels in a cell. The remaining $\mathbf{S - S_C}$ channels are shared by both originating calls and handover requests. A new call is blocked if number of available channels in the cell is less than or equal to S_C when the call is originated. A handoff attempt is unsuccessful if no channel is available in the target cell. Both the handoff calls and new calls are generated according to Poisson process with mean arrival rate λ_H and λ_O respectively. We define the sate Q_i of a cell such that total i calls

are in progress for the base station of a cell. If P_i is the steady state probability that the base station in sate Q_i, then the probabilities can be determined in the usual way for birth-death processes. The state transition diagram for this process is described as

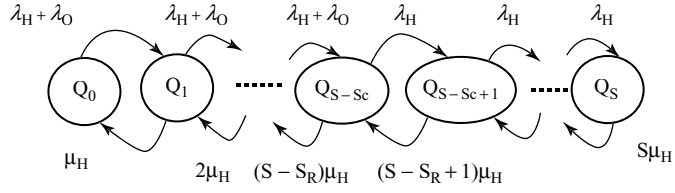


Fig. 3.21 *State transition diagram for the hand off queuing scheme I*

From Fig. 3.21, the steady state equations for P_i can be given as under

$$P_i = \begin{cases} \dfrac{\lambda_H + \lambda_0}{\mu_H} P_{i-1.} & \text{for } i = 1,2, \ldots\ldots, S - S_c \\ \dfrac{\lambda_H}{\mu_H} P_{i-1.} & \text{for } i = S - S_c + 1, \ldots\ldots S \end{cases}$$

The summation of all probabilities,

$$\sum_{i=1}^{\infty} P_i = 1$$

Finally, the steady state probability distribution $\{P_i\}$ is given as

$$P_i = \begin{cases} \dfrac{(\lambda_H + \lambda_0)^i}{i!\,\mu_H{}^i} P_0 & \text{for } i = 1,2, \ldots\ldots, S - S_c \\ \dfrac{(\lambda_H + \lambda_0)^{S-S_c} (\lambda_H)^{i-(S-S_c)}}{i!\,\mu_H{}^i} P_0 & \text{for } i = S - S_c + 1, \ldots\ldots S \end{cases}$$

where

$$P_0 = \left[\sum_{j=0}^{S-S_c} \frac{(\lambda_H + \lambda_0)^j}{j!\,\mu_H{}^j} + \sum_{j=S-S_c+1}^{S} \frac{(\lambda_H + \lambda_0)^{S-S_c} (\lambda_H)^{j-(S-S_c)}}{j!\,\mu_H{}^j} \right]^{-1}$$

The probability of blocking a new call is the sum of the probabilities that the state number of the base station is larger than or equal to $S - S_c$. Thus,

$$P_B = \sum_{i=S-S_c}^{S} P_i$$

The probability of handoff attempt failure or forced termination P_F is the probability that the state number of the base station is equal to S. Thus,

$$\mathbf{P_F = P_S}$$

Priority of Handoff Calls with Queuing Facility: Scheme II Same channel sharing method is used as of scheme I, but queuing facility for handoff calls is given. In this scheme, priority is given to handoff requests by assigning $\mathbf{S_C}$ channels exclusively for hand-off calls among the **S** channels in a cell. The remaining $\mathbf{S\text{-}S_C}$ channels are shared by both originating calls and handover requests as shown in Fig. 3.22. An originating call is blocked if the number of available channels in the cell is less than or equal to $\mathbf{S_C}$. If the BS finds all channels in the target cell is occupied, a **handoff request** is put in the queue. If a channel is released when the queue for hand-off requests is not empty, the channel is assigned to request on the top of the queue. In this scheme, the size of the queue is assumed to be infinite. If the received signal strength from the current BS falls below the receiver threshold level prior to the mobile being assigned a channel in the target cell, the call is forced to termination. The **First Input First Output** (FIFO) queuing strategy is used and infinite queue size at the BS is assumed. An originating call in the queue is deleted from the queue when it moves out of the cell before getting a channel. Also a hand-off request is deleted from the queue when it passes through the hand-off area before getting a new channel (i.e., forced termination) or the conversation is complete before passing through the hand-off area. A blocked handoff call can repeat trial handoffs until the received signal strength goes below the receiver threshold.

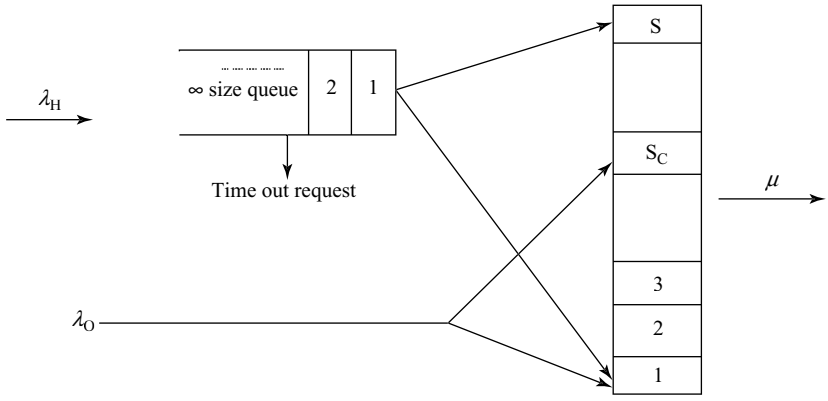


Fig. 3.22 *Hand-off call queuing: Scheme II*

The state transition diagram for this scheme is shown in Fig. 3.23.

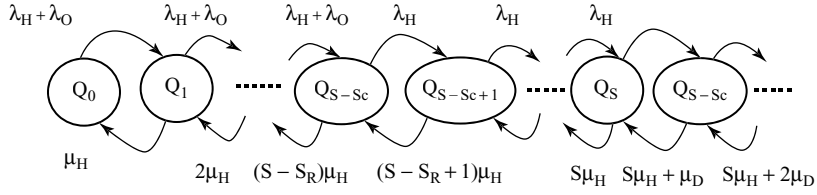


Fig. 3.23 *State transition diagram for priority hand-off call: Scheme II*

The time for which a mobile is in the hand-off area depends on the system parameters such as the speed and direction of the mobile travel and the cell size. This is called the dwell time or cell residence time of the mobile. As explained earlier, the dwell time is exponentially distributed with mean $1/\mu_D$. For those states whose state number 'i' is less than or equal to S, the state transition relation is the same as for scheme I. In the usual way for birth-death process, the probability distribution of state transition is given as

$$P_i = \begin{cases} \dfrac{(\lambda_H + \lambda_0)^i}{i!\,\mu_H{}^i} P_0 & \text{for} \quad 1 < i < S - S_c \\ \dfrac{(\lambda_H + \lambda_0)^{S-S_c}\,(\lambda_H)^{i-(S-S_c)}}{i!\,\mu_H{}^i} P_0 & \text{for} \quad S - S_c + 1 \le i \le S \\ \dfrac{(\lambda_H + \lambda_0)^{S-S_c}\,(\lambda_H)^{i-(S-S_c)}}{S!\,\mu_H{}^S \prod_{j=1}^{i-S}(S\,\mu_H + j\,\mu_D)} P_0 & \text{for} \quad i \ge S + 1 \end{cases}$$

where

$$P_0 = \left[\sum_{j=0}^{S-S_c} \frac{(\lambda_H + \lambda_0)^j}{j!\,\mu_H{}^j} + \sum_{j=S-S_c+1}^{S} \frac{(\lambda_H + \lambda_0)^{S-S_c}\,(\lambda_H)^{j-(S-S_c)}}{j!\,\mu_H{}^j} + \sum_{j=S+1}^{\infty} \frac{(\lambda_H + \lambda_0)^{S-S_c}\,(\lambda_H)^{j-(S-S_c)}}{S!\,\mu_H{}^S \prod_{k=1}^{j-S}(S\,\mu_H + k_{\mu_D})} \right]^{-1}$$

The blocking probability $\mathbf{P_B}$ is the sum of the probabilities that the state number of the base station is larger than or equal to $S - S_C$. Thus,

$$\mathbf{P_B} = \sum_{i=S-S_c}^{\infty} P_i$$

Summary

This chapter has dealt with the basic understanding of cellular communications, fundamentals, design issues related to frequency reuse, capacity expansion, etc. The main purpose of cellular mobile communication is to provide mobility to the users while roaming within the geographical regions. The radio frequency is very scarce. So the efficient use of it is very much required by the frequency reuse principle and method of capacity expansion such as sectoring of antenna, cell splitting. As the number of users increases day-by-day, capacity expansion is also needed by the use of increased frequency reuse factor. But problem arises due to co-channel interference and adjacent channel interference. Tradeoff is required for handling the problems of interference and design of cellular systems. As the size of the cell decreases due to smaller radio coverage of the cells to increase the system capacity, problem of handoff management arises. The resource management in cellular communications includes mobility management, call admission control, and channel allocation. This chapter highlights the fundamentals for all these topics. Moreover, this chapter describes the most important technique of prioritization of hand-off calls and performance parameters related to quality of service of the wireless cellular networks.

References

[1] Lee W.C.Y., *Mobile Communications Engineering: Theory and Applications*, 2nd edition, McGraw-Hill, New York, 1998.

[2] Rappaport T. S., *Wireless Communications: Principles and Practice*, Prentice Hall, PTR, 1996.

[3] Lee W.C.Y., Mobile Communication Design Fundamentals, John Wiley and Sons, Inc. 1993.

[4] Misra A. R., Fundamentals of Cellular Network Planning and optimization 2G/2.5G.3G …Evolution to 4G, John Wiley and Sons, England, 2004.

[5] Alencar M. S., V.C. da Rocha, Jr, *Communication Systems*, Springer, 2005

[6] Zander J., and S. L. Kim, *Radio Resource Management for Wireless Networks*, Artech House, Inc, 2001

[7] Linmartz M. G., Jean-Paul *Wireless Communication, The Interactive Multimedia* CD-ROM, Baltzer Science Publishers, Amsterdam, ISSN 1383 4231, Vol. 1 (1996), No.1

[8] Mark Jon. W., and Weihua Zhuang, *Wireless Communications: and Networking*, PHI, New Delhi, 2005.

[9] Hong D., and T. S. Rappaport, " Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Nonprioritized Handoff Procedure", IEEE Transactions on Vehicular Technology, Vol. VT-35, NO. 3. Aug 1986.

[10] Schiff L., "Traftic Capacity of Three Types of Common-user Mobile Radio Communication Systems," *IEEE Trans. Commun. Technol.,* vol. COM-18, pp. 12–21, Feb. 1970.

[11] Cox D. C., and D. O. Reudink, "Dynamic Channel Assignment in High- Capacity Mobile Communications Systems," *Bell Syst. Tech. J.,* Vol. *50,* pp. 1833–1857, 1971.

[12] Anderson L. G., "A Simulation Study of Some Dynamic Channel Assignment Algorithms in a High Capacity Mobile Telecommunications System," *IEEE Trans. Veh. Technol.,* Vol. VT-22, pp. 210–217, 1973.

[13] Kahwa T. J., and N. D. Georganas, "A Hybrid Channel Assignment Scheme in Large-scale Cellular-structure Mobile Communication Systems," *IEEE Trans. Commun.,* Vol. COM-26, pp. 432438, 1978.

[14] Cooper R. B., *Introduction to Queueing Theory,* 2nd Ed. New York: Elsevier North Holland, 1981.

[15] Wells J. D., "Cellular System Design Using the Expansion Cell Layout Method," IEEE Trans. Veh. Technol., vol. VT-33, May 1984.

Questions for Self-Test

3.1 Interference effect in cellular systems is a result of _____.
- a. The distance between areas
- b. The power of the transmitters
- c. The ratio of the distance between areas to the transmitter power of the areas
- d. The height of the antennas

3.2 Larger cells are more useful in _____.
- a. densely populated urban areas
- b. rural areas
- c. lightly populated urban areas
- d. mountainous areas

3.3 Cells are always hexagonal in shape
- a. True
- b. False

3.4 Frequency reuse is maximized by increasing the size of cells
- a. True
- b. False

3.5 Fixed wireless access is primarily a rural application
- a. True
- b. False

3.6 By using frequency reuse the system capacity can be increased without the use of high power transmitter
   a. True
   b. False

3.7 The mean signal strength for an arbitrary transmitter-receiver (Tx-Rx) separation is useful in estimating the radio coverage
   a. True
   b. False

3.8 Co-channel cells use the same set of frequencies
   a. False
   b. True

3.9 The co-channel interference are independent of base station transmitted power but on cell radius and co-channel distance
   a. True
   b. False

3.10 Co-channel reuse ratio or frequency reuse ratio (q) is related to _____.

3.11 The larger value of q (frequency reuse ratio) improves the transmission quality, whether a small value of q enhances system capacity.
   a. True
   b. False

3.12 The base station antennas are designed to achieve the desired coverage within a particular _____.

3.13 The design process of selecting and allocating channel groups for all of the cellular base station within a cellular system is known as frequency _____.

3.14 What is called radio cell? How does a cluster form?

3.15 What is the relationship between the cluster size and frequency reuse ratio? Give the different allowable cluster size in cellular system.

3.16 How is frequency reuse concept useful in cellular communication?

3.17 What is the role of S/I ratio in the design of cell cluster?

3.18 What are the important parameters that characterize frequency reuse?

3.19 The capacity expansion of cellular networks does not come without a price? What is the price to pay for capacity enhancement?

3.20 How are locations of co-channel cells determined in a cellular system, explain with pictorial representation.

3.21 What is called co-channel distance D? On what factors does D depend?

3.22 What is co-channel interference? How is signal to interference ratio (S/I) related with frequency reuse ratio q?

3.23 Determine the cell cluster size N with respect to two integers i and j associated in determining co-channel cells. With pictorial representation, first explain the meaning of the integers i and j.

3.24 Derive the expression for S/I ratio in a worst-case scenario of a mobile node within a cell? Repeat the same for 120^0 and 60^0 sectoring?

Ans. $(S/I)_{120}{}^{0} = \dfrac{R^{-\kappa}}{D^{-\kappa} + (D + 0.7\,R)^{-\kappa}}$ $(S/I)_{60}{}^{0} = R^{-\kappa} / D^{-\kappa.}$

3.25 Cell splitting is one of the methods of increasing capacity of cellular system.
   a. Discuss the method of cell splitting and show how it helps to increase capacity when a large cell of radius R is split into smaller cell of radius R/4.
   b. What is the price to pay for taking the advantage of increase capacity with cell splitting?

Ans. b. 16 x N small cells

3.26 When a cell with radius R is split to a new cell of radius R/2, what will be the base station transmit power for the two cases. Find the transmit power ratio in dB. Does cell splitting increase capacity of the system?

Ans. Power in dB = 10 $\kappa \log_{10} 2$

3.27 How is received power at the mobile station related with distance and path loss exponent?

3.28 Discuss the concept of macro, micro and pico cells. When do they become useful?

3.29 Discuss the various method of increasing capacity in a cellular wireless system? High light the pros and cons of those methods.

3.30 A cellular system with 25 MHz total bandwidth is allocated for duplex communication. Each simplex channel is 30 kHz. The frequency reuse factor for the system is 7. Find

a. The total number of duplex channels

b. The total number of channels per cell

Ans. a. 416, b. 59

3.31 A Cellular system has a total of 500 duplex voice channels without frequency reuse and uses omnidirectional antenna. The service area is 1000km^2 and area of each cell is 5 km^2. The required signal-to-co-channel interference ratio is 18 dB. Consider the path loss exponent κ equal to 3, and 4 respectively. Determine

a. the cell cluster size,

b. the number of cell clusters in the service area, and

c. the maximum number of users in service at any instant.

Discuss the effects of the path loss exponent on the frequency reuse. (Consider the first tier neighbor cells).

Ans. $\kappa = 3$, a. 19 b. 10 c. 5000, $\kappa = 4$ a. 7 b. 28 c. 14000

3.32 Now in Problem 3.26, 120^0 sectoring and 60^0 sectoring are done. Find the optimal value of frequency reuse factor N for these sectoring if signal to noise ratio is taken as 18 dB in the worst-case scenario. Which sectoring will be suitable for this system if $\kappa = 4$?

Ans. N=7 and N=4, 60° sectoring

3.33 What are the different channel allocation schemes used in cellular communications? In which situations each of these schemes are suitable?

3.34 Explain the mobility management for cellular communication network?

3.35 What are called in-session and out-of-session mobility steps?

3.36 Define hand-off process? What are the different types of handoff used in cellular communication?

3.37 What are the different parameters generally defined for the performance measure of a cellular communication system?

3.38 How does prioritization of hand-off calls over originating calls help for reducing blocking and drop calls?

Characteristics of Wireless Channels and Propagation Path Loss Models

4

Introduction

In wireless communications, the radio wave propagation medium largely affects the transmitted signal in between the path of transmitter and receiver. As the signal propagates through wireless channel, it experiences random fluctuations in time if the transmitter, receiver, or surrounding objects are moving, due to changing reflections and attenuation. Thus, the characteristics of the channel appear to change randomly with time. Unless, the channel response is known, it is difficult to design reliable systems with certain reliability and guaranteed quality.

The wireless propagation channel contains large obstacles scattering radio signal. A receiver may get many multipath signal components from various scattering objects those may interfere constructively or destructively depending on the phases of the multipath components and causing the phenomenon like fading. Further, the speed of the mobile device affects the signal level that also causes fading. Unlike free space propagation, scatterer introduces channel impairments like fading, multipath, delay spread, Doppler spread, etc. The signal fading may be frequency selective or non-frequency selective depending on the channel characteristics parameters. Thus, wireless channel is time variant in nature.

In this chapter, the wireless channel characteristics along with the radio propagation path loss models in wireless environments are discussed. The variation of received signal power over distances due to path loss and shadowing are characterized. Path loss is caused by dissipation of the power radiated by the transmitter as well as effects of propagation channel. Obstacles between the transmitter and receiver cause shadowing that attenuate received power through reflection, absorption, diffraction and scattering. Variation due to path loss occurs over large distance (few hundreds meters), whereas variation due to shadowing occurs over distances proportional to the length of the obstructing object (10–100 meters).

4.1 MULTIPATH PROPAGATION MECHANISMS

There are several mechanisms for creating multiple propagation paths. These are given below.

1. **Reflection:** Reflection occurs when a propagating electromagnetic waves falling on objects with smooth surface and large dimensions compared to wavelength.
2. **Diffraction:** It is caused by obstructions with sharp irregularities in the path of the radio signal. Diffraction occurs when the radio path between the transmitter and receiver is obstructed by a dense body with large dimensions compared to λ, causing secondary waves to be formed behind the obstructing body. Diffraction is a phenomenon that accounts for RF energy traveling from transmitter to receiver without a line-of-sight path between the two. It is often termed shadowing because the diffracted field can reach the receiver even when shadowed by an impenetrable obstruction.
3. **Refraction:** The propagation path is diverted because of the difference in the electrical properties of the medium.
4. **Scattering:** This is caused due to obstacles in signal path with much smaller dimensions than signal wavelength. These objects could be water droplets, clouds, or insects for example. Scattering causes the reflected energy to spread out (scatter) in all directions. In an urban environment, typical signal obstructions that yield scattering are lampposts, street signs, and foliage.
5. **Multipath:** Signals reflected through buildings, mountains, trees, open ground, etc.

6. **Fading:** This includes slow and fast (Rayleigh model), Doppler Shift. Figure 4.1 shows the various mechanism of the multipath phenomena for radio waves.

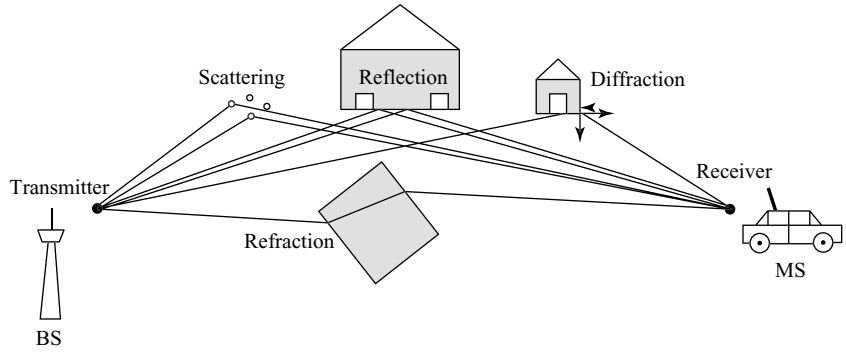


Fig. 4.1 *Different mechanisms creating multipath*

These various mechanisms give rise to alternate propagation paths such that the received signal is a composite of numerous replicas all differing in phase, amplitude, and in time delay. Thus, the multipath signal amplitudes, phases, and time delays become random variables.

In order to better understand the performance of a wireless signal propagating in a typical outdoor propagation environment, it is necessary to define some terms those are commonly used to describe properties or characteristics of the channel.

Fading is a term used to describe the fluctuations in a received signal as a result of multipath components. Several replicas of the signal arrive at the receiver, having traversed different propagation paths, adding constructively and destructively. The fading can be defined as fast or slow fading. Additionally, fading can be defined as *flat* or *frequency selective fading*.

4.2 MULTIPATH EFFECTS IN MOBILE COMMUNICATION

Radio propagation in a mobile communication environment, both indoor and outdoor, is a complex phenomenon and suffers dispersion from various structures like tall buildings, etc. The signal offered to the receiver contains not only a direct line-of-sight (LOS) radio wave, but also a large number of reflected radio waves (as shown in Fig. 4.2). These reflected waves interfere with the direct wave, which causes significant degradation of the performance of the wireless channel. A wireless channel has to be designed in such way that the adverse effect of these reflections is minimized.

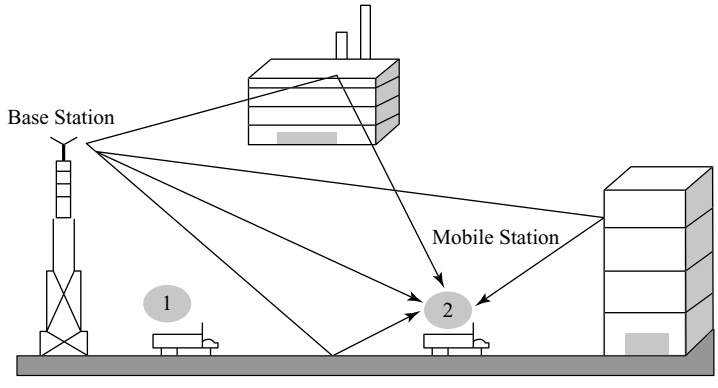


Fig. 4.2 *Multipath propagation*

A single pulse transmitted over a multipath channel, will appear as a pulse train. Each pulse in the train contains the distinct multipath components associated with a distinct scatterer or cluster of scatterer along with the LOS component. Because multiple propagation paths have different delays, characteristic of a multipath channel is the time delay spread to the received signal. The delay spread is defined as the time delay between the first received signal component and the last received signal component associated with a single transmitted pulse. The delay spread is related to duration of the pulse *T*, which is the inverse of the signal bandwidth. If the delay spread is small compared to *T*, then little time spreading in the received signal will occur, otherwise if time delay is large, significant delay dispersion to the received signal will occur resulting distortion to the received signal.

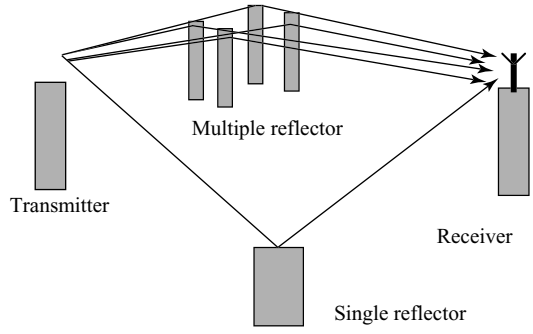


Fig. 4.3 *Single and multiple reflectors causing multipath*

Another multipath time variation of the signal occurs because of the motion of the transmitter or receiver and therefore the location of reflectors in the transmission path. This is again changed over the time. Thus, if a pulse is repeatedly transmitted from a moving transmitter, varied amplitudes, delays, and the number of multipath components corresponding to each pulse will be observed. These changes occur over a much larger time scale than the fading due to constructive and destructive addition of multipath components associated with a fixed set of scatterers.

Consider that an antenna transmits two narrow pulses at a fixed distant mobile unit, as shown in Fig. 4.4.

The first received pulse is the LOS signal. But in practical cases, magnitude of received pulse changes with some distortion due to propagation attenuation. With the increasing speed of mobile unit, the LOS attenuation increases. The multiple secondary pulses are due to scattering, reflection and diffraction. The multiple delayed pulses are the source of error that reaches at the same time at the receiver with the primary pulse. So, recovery of the original transmission becomes difficult. The number, magnitude and timing of the secondary pulse vary with respect to the movement of mobile unit in different location having varied obstacles.

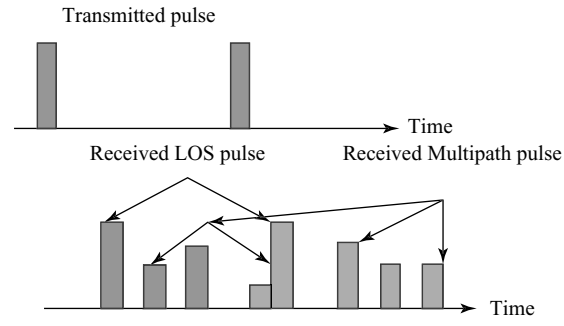


Fig. 4.4 *Two pulses in time variant path*

4.2.1 Types of Fading

Fading as caused by multipath, due to reflections on buildings and natural obstacles in outdoor environment, or walls, roof, floors in indoor environments, produce series of dips in the received signal spectrum. Depending on the signal parameters such as *bandwidth*, *duration of symbol* and the type of channel parameters causing rms delay spread due to time dispersion of multipath propagation or Doppler spread due to velocity of the mobiles, the transmitted signal may undergo different types of fading. The time dispersion and frequency dispersion mechanisms in a mobile radio channel are two independent phenomena generate four different distinct fading scenarios. The multipath delay spread causes **time dispersion and frequency selective fading** and the Doppler spread leads to **frequency dispersion and time selective spreading.** The transmitted signal undergoes either flat or frequency selective fading when time dispersions occur due to multipath propagation.

Frequency Selective and Non-frequency Selective (Flat fading) Fading If the bandwidth of the transmitted signal is less than the bandwidth of the mobile radio channel (**Bs < Bc**) and also Delay spread < Symbol period, then all the frequency components of the signal would roughly undergo the same degree of fading. The channel is then classified as **frequency non-selective** or **flat fading**.

The received signal has a constant gain and linear phase response over a bandwidth greater than the bandwidth of the transmitted signal. Though the spectral characteristics of the transmitted signal is preserved at the receiver, but signal strength of the received signal is varied with time due to the fluctuations in channel gain. Thus flat fading channel is **amplitude varying**. Flat fading scenario is illustrated in Fig. 4.5.

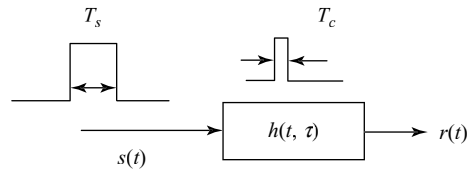


Fig. 4.5 *Flat fading scenario*

In this case, delays between different paths are relatively small with respect to the symbol duration. We can assume that we would receive only one copy of the signal, whose gain and phase are actually determined by the superposition of all those copies that come within coherence time. When the signals with identical delays recombined produces distortion known as **narrowband fading**.

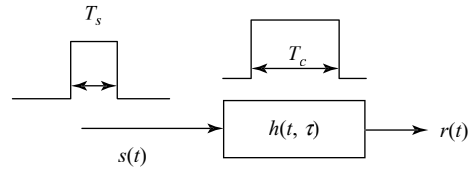


Fig. 4.6 *Frequency selective fading scenario*

On the other hand, **frequency selective fading** occurs if the received signal undergone time dispersion within a channel under the condition that **bandwidth of the transmitted signal is greater than the bandwidth of the channel** ($\mathbf{B_s > B_{ch}}$). Also the delay spread due to multipath propagation is greater than the transmitted symbol period ($\tau > T_s$). The received signal in frequency selective fading is the distorted version of the transmitted signal that includes multiple version of the transmitted waveform attenuated and delayed in time as shown in Fig. 4.7. This frequency selective fading is the cause

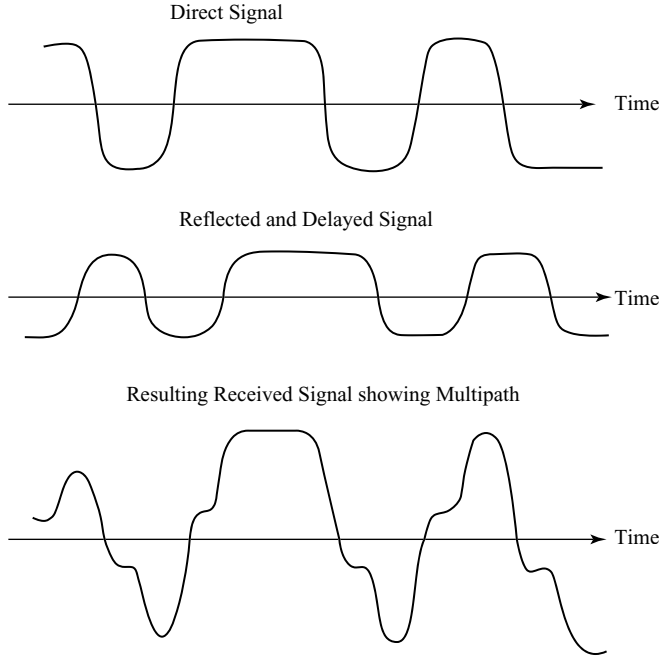


Fig. 4.7 *Effect of ISI on received signal*

of **Inter Symbol Interference** (ISI) of the channel. In the frequency domain representation of the received signal, it can be observed that certain component of the received signals have greater gain than the other. Frequency selective fading is also known as **wideband fading** because of the recombination with different delays. Modeling of frequency selective fading is more difficult than flat fading channel.

In short, it can be said that for **flat fading** fluctuations of the received signal for all frequencies undergone same proportion simultaneously. Selective fading on the other hand affects different spectral components of the radio signal received unequally.

Fast and Slow Fading A channel can be again considered as the **fast fading or slow fading channel**. These are the **fading effects due to the Doppler spread**. In a slow fading channel, the channel impulse response changes at a rate much slower than the rate of transmitted baseband signal and the channel is almost assumed to be static over one or several reciprocal bandwidth intervals. In the fast fading, impulse response changes rapidly within the one transmitted symbol duration. This is because the coherence time of the channel smaller than the symbol period of transmitted signal. This is the cause of frequency distortion or **time selective fading**. Higher user mobility produces higher Doppler spread, means higher time selective channel. Coherence time is given as the inverse of Doppler spread ($t_c \approx 1/f_d$).

If symbol duration/ coherence time (T_s/t_c) ≤1, the channel is **more time selective, and result is fast fading.** If (T_s/t_c) >>1, the channel is less time selective and result is **slow fading.**

Slow fading is characterized as the mean value of the set of signal fluctuations whereas **fast fading** is characterized by the variation of signal values around its average value.

At the typical operating frequency of cellular communications (900 MHz), the wavelength is of the order of 0.33 m. In an urban environment, the amplitude of the received signal may get a change of 20 to 30 dB over a short distance. This rapid change of fading is known as *fast fading*. When the average received signal changes slowly over a long distance as mobile traverses in the urban environment, the fading is known to be *slow fading.* Fig. 4.8 is the typical representation of fast and slow fading.

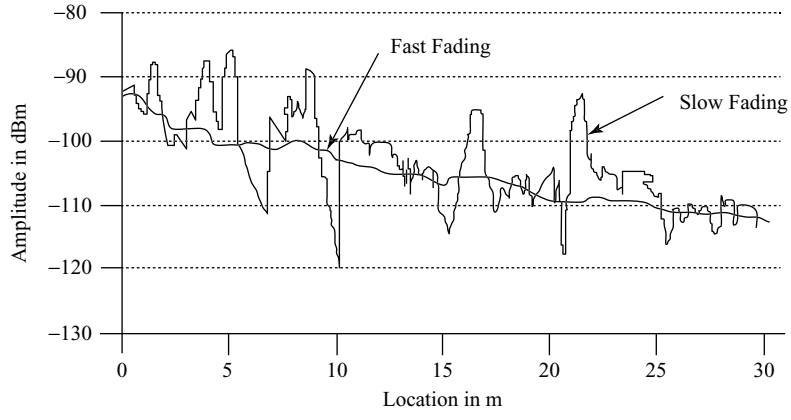


Fig. 4.8 *Fast and slow fading in urban area*

Large and Small-Scale Fading Large scale fading is affected by prominent terrain contours like hills, forests, large buildings, etc. between the transmitter and receiver. The receiver is represented often as being "Shadowed" by such obstacles. The statistics of large-scale fading provide a way of computing an estimate of path loss as a function of distance. This is described in terms of a mean-path loss (n^{th}-power law) and a log-normally distributed variation about the mean. The mean power averaged over about 10^{th} of wavelengths and the fluctuation occurs over a large distance (few hundreds of wavelengths) and is considered as **large-scale fading**.

Small-scale fading refers to the dramatic changes in signal amplitude and phase that can be experienced as a result of small changes (as small as a half-wavelength) in the spatial separation between a receiver and transmitter. Small-scale fading manifests itself in two mechanisms, namely, timespreading of the signal (or signal dispersion) and time-variant behavior of the channel. On a very short distance scale (comparable to one wavelength), if the received power fluctuates around a local mean value then the fading is said to be **small-scale.** This is occurred due to the interference between the multiple components of the transmitted signal from different paths.

Small-scale fading is also called *Rayleigh fading* because if the multiple reflective paths are large in number and there is no line-of-sight signal component, the envelope of the received signal is statistically described by a Rayleigh probability distribution function. When there is a dominant non-fading signal component present, such as a line-of-sight propagation path, the small scale fading envelope is described by a Rician probability distribution function. A mobile radio roaming over a large area must process signals those experience both types of fading: small-scale fading superimposed on large-scale fading.

Table 4.1 *Different propagation mechanisms*

Cause	Effect
Multipath propagation	Fast fading, Delay Spread (Time dispersion)
Motion	Doppler Shift (frequency dispersion)
Shading	Slow fading
Signal Attenuation	Path Loss

4.3 CHARACTERIZATION OF THE CHANNEL

Channel characterization is an important topic in the investigation of mobile channel. Modeling the attenuation as a function of frequency, location and distance; time variation measurement of amplitude and phase; use of correlation for improved result; use of diversity techniques and to establish the channel impulse response to determine the delay spread are some of the usual way to address the channel characterization. Although channel fading is experienced as an unpredictable, stochastic phenomenon, powerful models have been developed that can accurately predict system performance. In this section, multipath channel is modeled due to the random time varying impulse response.

Linear Time Invariant (LTI) system does not have frequency component different from those of the input signal, thus no frequency shifts occur. But both nonlinear and time-varying systems introduce new frequency components other than those existing in the input signal. For wireless propagation environment, due to the mobility of mobile users and/or the surrounding scatterers, the channel is linear but time variant and is defined as Linear Time Variant (LTV) channel.

4.3.1 Time Varying Channel Impulse Response

Consider there are N number of scatterer or reflected objects in between the transmitter and receiver causing N multipaths received signals. We can express the received signal phasor as the sum of all possible multipath components within the receiver. Consider a narrowband signal $S(t)$, is transmitted at frequency f over the wireless channel,

$$S(t) = \text{Real of } \left\{ s(t)\, e^{j2\pi ft} \right\} \tag{4.1}$$

Where $s(t)$ is the complex envelop of the signal.

$$\text{The received signal } R(t) = \text{Real of} \left\{ \sum_{n=1}^{N} a_n(t) \times (t - \tau_n(t))^{ej2\pi f(t-\tau_n(t))} \right. \tag{4.2}$$

where $a_n(t)$ is the amplitude variation of the different multipath received signal, $\tau_n(t)$ is the different time delay associated with multipath.

$$R(t) = \text{Real of} \left\{ \sum_{n=1}^{N} r(t)\, e^{j2\pi ft} \right.$$

$$r(t) = \sum_{n=1}^{N} a_n(t)\, x\, (t - \tau_n(t)) \tag{4.3}$$

The received signal is the convolution of the input transmitted signal and the channel impulse response.

$$r(t) = h(t)^* s(t) = \int_{-\infty}^{\infty} s(t-\tau)\, h(\tau)\, \text{dt} \tag{4.4}$$

Thus the channel can be characterized by the channel impulse response at the baseband signal. As the channel is time varying in nature, the impulse response depends on the time instant the input signal applied. Let $\delta(t_1)$ and $\delta(t_2)$ are the inputs to the system for getting channel responses at instant t_1 and t_2 respectively. For time varying channel, channel response $h_1(t_1)$ at time t_1 is not same with the channel response $h_2(t_2)$ at time t_2. As a result, to characterize an LTV channel, two parameters are used to define impulse response and is denoted by $h(t, \tau)$, where τ represents the propagation delay. The received signal for LTV is given as

$$r(t) = h(t)^* s(t) = \int_{-\infty}^{\infty} s(t-\tau)\, h(t,\tau)\, d\tau \tag{4.5}$$

where,

$$h(t,\tau) = \sum_{n=0} a_n\, e^{j\varphi n} \delta(t - \tau_n(t)) = \sum_{n=0} a_n\, e^{j\varphi n} \delta(t - n\Delta\tau) \tag{4.6}$$

a_n = time varying amplitude of path n, φ_n = time varying phase of path n, that may include Doppler effect, $\tau_n(t)$ = time varying delay of path n.

The complex-valued channel impulse response can be given in terms of uniformly spaced channel sampled as shown in Fig. 4.9.

The received signal,

$$r(t) = \sum_{n=0}^{N-1} a_n e^{j\varphi n} s(t - \tau_n) \tag{4.7}$$

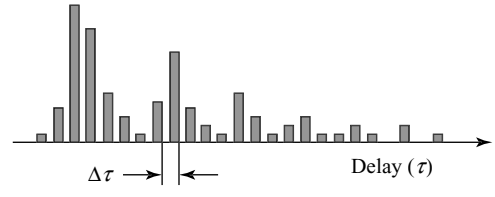


Fig. 4.9 *Sampled value of channel impulse response*

The magnitude plot of an impulse response magnitude, $|h(t, \tau)|$, is shown in Fig. 4.10.

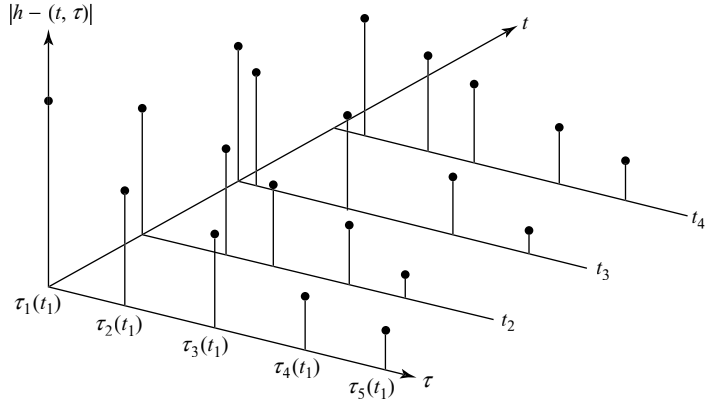


Fig. 4.10 *Magnitude of channel impulse response at different times*

The **time variant transfer function** of the LTV channel $H(f, t)$, is the Fourier transform of the time varying channel impulse response $\boldsymbol{h(t,\tau)}$ **with respect to** τ.

$$H(f,t) = \textbf{FT of } h(t,\tau) = \int_{-\infty}^{\infty} h(t,\tau).e^{-j2\pi f\tau} d\tau \tag{4.8}$$

$$h(t,\tau) = \textbf{IFT of } H(f,t) = \int_{-\infty}^{\infty} H(f,t).e^{+j2\pi f\tau} df \tag{4.9}$$

Thus $\boldsymbol{H(f, t)}$ and $\boldsymbol{h(t,\tau)}$ are the Fourier transform pair.

$$\boxed{h(t,\gamma) \underset{IFT}{\overset{FT}{\rightleftharpoons}} H(f,t)}$$

The received signal in frequency domain is given as,

$$R(f,t) = \boldsymbol{H(f,t)}S(f), \text{ where } S(f) = \text{FT of } s(t) \tag{4.10}$$

Let us define the **Doppler Spread Function** (DSF) as $H(f,\gamma)$ which is the channel gain associated with Doppler shift to the input signal at frequency f and γ is the frequency shift due to Doppler effect. The received signal at frequency domain in the LTV channel is given as

$$R(f) = \int_{-\infty}^{\infty} X(f-\gamma)\, H(f-\gamma) d\,\gamma, \text{ where } \gamma \text{ is the Doppler frequency shift}$$

The relationship between the **Doppler Spread Function** (DSF) and the channel time variant transfer function is given as,

$$H(f,\gamma) = \textbf{FT of } H(f,t) = \int_{-\infty}^{\infty} H(f,t)e^{-j2\pi\gamma t} dt \tag{4.11}$$

and $H(f, t)$ is the inverse Fourier transform of $H(f, \gamma)$

$$\boxed{H(f,t) \underset{IFT}{\overset{FT}{\rightleftharpoons}} H(f,\gamma)}$$

For LTV channel there is another function defined as Delay-Doppler Spread Function which is the FT of channel impulse response $h(t,\tau)$ with respect to t.
Thus **Delay Doppler Spread** (DDS) is expressed as

$$H(\tau,\gamma) = \textbf{FT of } h(t,\tau) = \int_{-\infty}^{\infty} h(t,\tau)e^{-j2\pi\gamma t} dt \tag{4.12}$$

The Doppler spread in frequency (f_d) is the inverse of channel coherence time (t_c),

$$t_c = 1/f_d \tag{4.13}$$

The channel coherence time should be much larger than the symbol duration of the transmitted signal to avoid fast fading.

4.4 MODELS FOR MULTIPATH RECEPTION

The effect of multipath reception produces rapid fluctuations of the signal amplitude and phase for a fast moving user, whereas dispersion and inter symbol interference for wideband digital signal. For a stationary

user of a narrowband system, good reception at some locations and frequencies may occur; while poor reception at other locations and frequencies may occur due to multipath fading.

In designing a communication system, the knowledge of the multipath fading effects and noise on the mobile channel is needed. The simplest channel model is the additive white Gaussian noise (AWGN) channel. But for mobile wireless transmission with varied environment, AWGN is not a good design solution for mobile channel.

Though the complete description of the mobile communication channel would be very complex and must take into account the effect of fading, multipath as well as interference, many models have been proposed to represent the envelop of the received signal.

4.4.1 Fast Fading Model

Assume a large number of randomly distributed scattering objects N, we can assume that the phases ϕ_n are uniformly distributed. We can express the time-domain version of the received signal as

$$R(t) = \sum_{n=1}^{N} a_n \cos(\omega_0 t + \phi_n) = \sum_{n=1}^{N} a_n \cos(\omega_0 t) \cos(\phi_n) - \sum_{n=1}^{N} a_n \sin(\omega_0 t) \sin(\phi_n) \qquad (4.14)$$

a_n, ϕ_n are the random amplitude and phase variation of the n th path respectively. Further, simplification can be done by considering,

$$X = \sum_{n}^{N} a_n \cos(\phi_n) \text{ and } Y = \sum_{n}^{N} a_n \sin(\phi_n)$$

$$R(t) = X \cos(\omega_0 t) - Y \sin(\omega_0 t) = A \cos(\omega_0 t + \phi)$$

where $r = \sqrt{X^2 + Y^2}$ envelop of the received signal and $\phi = \tan^{-1}$ (Y/X).

In the limit, as $N \to \infty$, the Central Limit Theorem dictates that X and Y follow the Gaussian random variables with mean zero and variance σ^2. The phase ϕ can also be modeled as a uniform distribution such that, $p(\phi) = 1/2\pi$ for $0 \leq \phi \geq 1/2\pi$. The envelope of the received signal 'r' follows the Rayleigh distribution.

Rayleigh distribution [5,10] is a well-known statistics for the amplitude modeling of radio signal in a fading environment. This distribution represents the effects of the amplitude of many signals, reflected, and refracted reaching to the receiver in a situation when there is no prevailing component or direction (Marcelo *et.al*, 2005). The well-known Rayleigh probability density function is given by the equation,

$$p(r) = r / \sigma^2 e - \frac{r^2}{2\sigma^2} \text{u(r)} \qquad (4.15)$$

With average (mean value), $E[r] = \sigma(\sqrt{\pi/2})$ and variance $V[r] = 2\sigma^2 (1 - \pi/4) = 0.429\sigma^2$, Mean square value = $2\sigma^2$. The location of maximum occurs {max. (pdf (r))} at $r = \sigma$. The associated phase distribution is uniform over the interval $(0,2\pi)$. The cumulative distribution function cdf (r) is defined as the probability that the realization of the random variable has a value smaller than r.

$$\text{cdf}(r) = \int_{-\infty}^{r} \text{pdf}(r)\text{dr} = 1 - \exp(-r^2 / 2\sigma^2 \simeq r^2 / 2\sigma^2 \text{ (for small values of } r) \qquad (4.16)$$

Rayleigh fading is caused by multiple receptions from various directions. The mobile antenna receives a large number, say N, reflected and scattered waves. Because of wave cancellation effects, the instantaneous received power seen by a mobile unit becomes a random variable, dependent on the location of the antenna.

Figure 4.11 shows the Rayleigh distribution function for specific σ values. The probability distribution is such that the area under the curve will be same for different σ values.

It can be shown that if the received signal envelop is Rayleigh distribution, then the instantaneous power follows exponential distribution. Thus pdf of power is given as

$$p(\gamma) = (1/\gamma_0)\, e^{-\gamma/\gamma_0} \tag{4.17}$$

where $\gamma_0 = E[\gamma] = \int_0^\infty \gamma\, p(\gamma)\, d\gamma = 2\sigma^2$. is the average expected energy of the received signal. To reach this point, it is assumed that all the NLOS components statistically have same energy.

Let the minimum detectable threshold power in a receiver is P_{th}. This power level includes the receiver noise floor, noise figure, and detector thresholds. If the received power falls below the threshold, the receiver goes

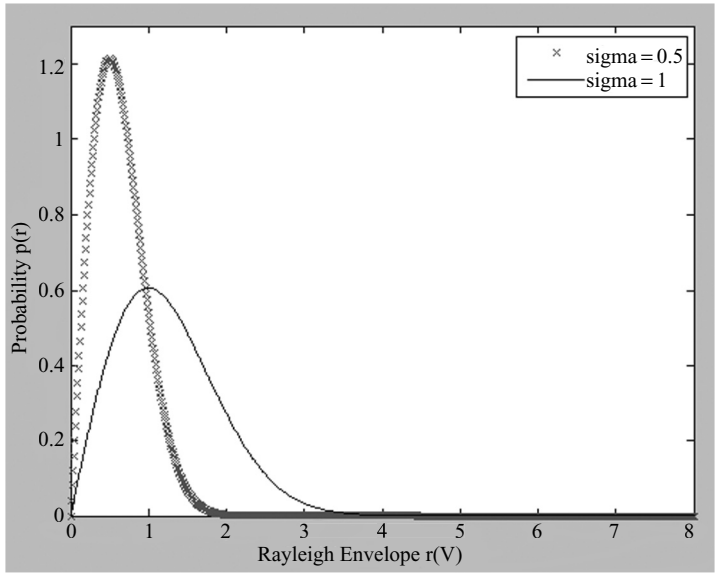


Figure 4.11 *Rayleigh distribution function for different* σ

into "outage" because the backend signal-to-noise ratio is insufficient. A new term" *outage probability"* is defined with the probability that the received power is too small for detection and is given by,

$$P(p \le P_{th}) = \int_0^{P_{th}} (1/\gamma_0)\, e^{-\gamma/\gamma_0}\, d\gamma, \text{ where } \gamma_0 = 2\sigma^2 \tag{4.18}$$

4.4.2 Multipath with Direct Component

In the previous modeling, only N number of multipath received signal components are considered without considering any direct LOS signal. In this section we will consider the modeling of multipath signal in presence of direct signal as shown in Fig. 4.12. This situation will be different when there is a component of strong signal due to some other phenomena other than reflections from the surrounding buildings, etc., i.e., there is a direct LOS scattering signal along with the NLOS signals. LOS scattering has a special component from the direct path and is modeled by Rician distribution instead of Rayleigh. In Rician fading a strong dominant component is present. This dominant component can, for instance, be the line-of-sight wave. Rician models also consider that the dominant wave can be a phasor sum of two or more dominant signals, e.g., the line-of-sight, plus a ground reflection. This combined signal is then mostly treated as a deterministic process.

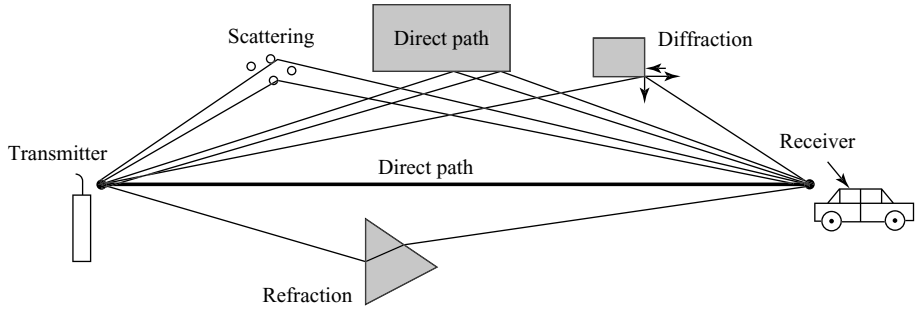


Fig. 4.12 *Multipath with direct LOS component*

The received signal is the sum of all multipath components and the LOS component.

$$R(t) = \sum_{n=1}^{N} a_n \cos(\omega_0 t + \phi_n) + D\cos(\omega_0 t) \tag{4.19}$$

$$N = \left(D + \sum_{n=1}^{N} a_n \cos(\omega_0 t) \right) \cos(\phi_n) - \sum_{n=1}^{N} a_n \sin(\omega_0 t) \sin(\phi_n) \tag{4.20}$$

Considering $X = D + \sum_{n}^{N} a_n \cos(\phi_n)$ and $Y = \sum_{n}^{N} a_n \sin(\phi_n)$

The random variable X is Gaussian with mean of D and standard deviation of σ. Random variable Y is Gaussian with zero mean and standard deviation of σ. The modified expression for Rician fading is given as follows.

$$r_{Ric}(t) = D + X(t) + jY(t)$$

Again $X(t)$ and $Y(t)$ are independent Gaussian statistical processes. The distribution function of the received envelop is the combination of Rayleigh distribution multiplied by some factor containing non-centrality parameter and a zero-order modified Bessel function of first kind. So, The probability density function for the envelope for Rician distribution is,

$$p_{RiC}(r) = J_0(r\,D/\sigma^2)(r/\sigma^2)e^{-(r^2+D^2)/2\sigma^2} \qquad r \geq 0,\ D \geq 0 \tag{4.21}$$

where J_0 is the zero order modified Bessel function, D denotes the dominant amplitude of constant signal. This distribution is known as *Rician distribution*. One important parameter for Rician channel fading is K factor that is defined as

$$K = D^2/2\sigma^2 \tag{4.22}$$

= Power of LOS component / Total power of all other scattered components

When K is infinity, only the LOS dominant component matters and there is no fading. On the other hand, when K is 0, the Rician distribution approaches to Rayleigh distribution. Stronger the LOS component, the rare the chances of deep fade. K is expressed in dB as

$$K(dB) = 10\log_{10}(D^2/2\sigma^2) \tag{4.23}$$

Figure 4.13 shows the Rician distribution for different values of K factor.

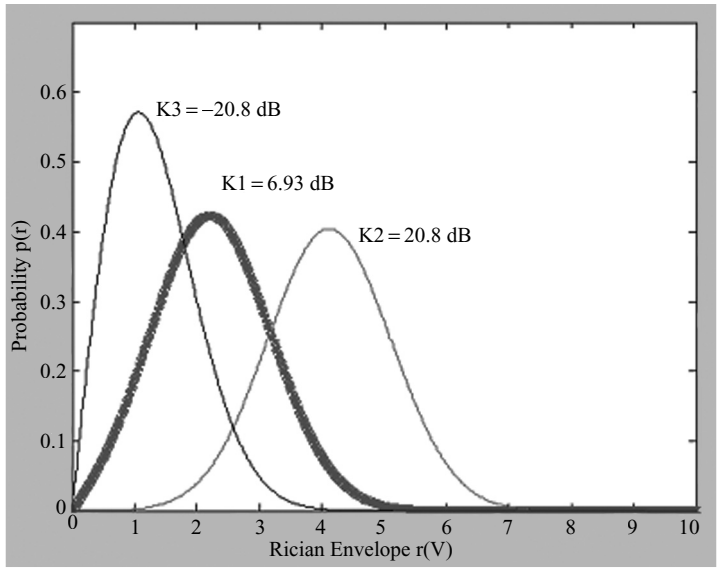


Figure 4.13 *Rician distribution for different K values*

It is seen that the knowledge of fading statistics is very important for designing the wireless channel. At the receiver the field strength is a random variable in a fading environment and does not guarantee successful communications at all times. There is an important point to know what will be the mean power to have successful communication in y% for a given field strength. This can be defined as the **Fading Margin**. Fading margin can be obtained from the definition of cdf (cumulative distribution function) that represents the probability that a certain field strength level is not exceeded [16]. Thus for y% of successful communication,

$$y = \text{cdf}(X_{\min}) = 1 - \exp(-x^2/2\sigma^2) \approx {x_{\min}}^2/2\sigma^2,$$

So, the fading margin is

$$2\sigma^2 = {x^2}_{\min}/Y \tag{4.24}$$

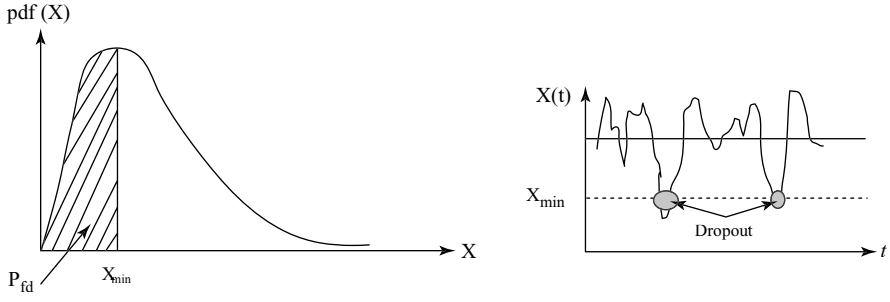


Figure 4.14 *Dropout probability (P_{fd})*

4.4.3 Effect of Velocity of the Mobile: Doppler Shift

With respect to Fig. 4.15, let the plane incident waves on a mobile unit moving along x direction with a velocity v.

Every wave incident on the mobile is subjected to Doppler shift due to the motion of the mobile receiver and arrives at the same time at the receiver. As the vehicle moves at a constant velocity as in Fig. 4.15, many factors change with time. The angles (α_n) of each multipath signal are time dependent. Each multipath experiences a different Doppler shift because the angle of scattering with respect to the moving vehicle is different for each scattering object. Also, the overall phase shift (θ_n) changes with time because the propagation delays are changing.

Thus the channel is linear but time variant. As a result the wireless channel introduces a frequency shift to the transmitted signal. The Doppler shift in Hertz for n^{th} wave incident at an arbitrary angle α_n is given as

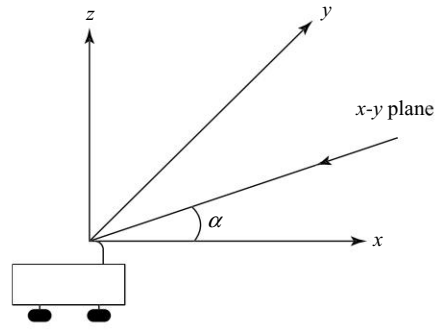


Fig. 4.15 *Plane wave arriving at an angle* α

$$f_n = f_d \cos \alpha_n = (v/\lambda) \cos \alpha_n, \text{ where } \lambda \text{ is the incident wavelength}$$

Considering the Doppler frequency shifts f_n, the received signal $r(t)$ is the sum of many reflected signal of N components and can be expressed as

$$r(t) = \sum_0^N a_n \cos(2\pi f_n t + \phi_n) \cos(\omega_0 t) - \sum_0^N a_n \sin(2\pi f_n t + \phi_n) \sin(\omega_0 t) \tag{4.25}$$

$$= \sum_0^N a_n \cos(2\pi f_d \cos(\alpha_n t + \varphi_n) \cos(\omega_0 t) - \sum_0^N a_n \sin(2\pi f_d \cos(\alpha_n) t + \varphi_n) \sin(\omega_0 t)$$

where there are three random variables, a_n, φ_n and α_n. The carrier frequency ω_0 is large compared to Doppler shift and is considered as narrowband operation. When N is very large, the envelop of the received signal is approximated with Gaussian random variables and the phase angle θ_n is assumed to be uniformly distributed over $0 \le \varphi_n$ and $\alpha_n \le 2\pi$. Thus, when superimposing N, statistically independent random variables, none of which is dominant, the associated probability distribution function follows the normal distribution for large N. In general, the envelop of the received signal r follows the Rayleigh distribution and is given as

$$r = \sqrt{X^2 + Y^2},$$

where $X = \sum_0^N a_n \cos(2\pi f_d \cos(\alpha_n) t + \varphi_n)$

$$Y = \sum_0^N a_n \sin(2\pi f_d \cos(\alpha_n) t + \varphi_n)$$

The amplitude of the received signal is normalized to get the ensemble average of the a_n 's equal to 1, i.e.,

$$\sum_{n=1}^N \overline{a_n^2} = 1 \tag{4.26}$$

As $X(t)$ and $Y(t)$ are Gaussian random processes, have uncorrelated zero mean Gaussian random variables X and Y at any time t with equal variance is

$$\overline{x^2} = \overline{y^2} = \frac{1}{2} \sum_{n=1}^N \overline{a_n^2} \tag{4.27}$$

So, the envelop of the received signal has a Rayleigh distribution with variance given above. Figure 4.16 shows the Rayleigh fading component due to multipath scattering.

The superposition of Doppler-shifted carrier waves leads to a **fluctuation** of the carrier amplitude and phase. This means that the received signal is amplitude and phase modulated by the channel. The effect of the Doppler shift of the transmitted signal is the *spectral broadening in frequency domain.*

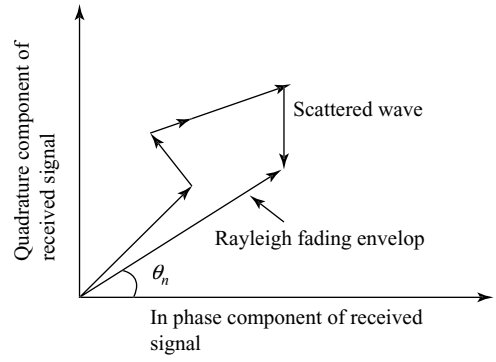


Fig. 4.16 *Rayleigh fading component in multipath scattering*

In case of a presence of direct LOS scattering signal along with the NLOS signals, Rayleigh fading component will be changed to Rician fading component as shown in Fig. 4.17. In Rician fading, a strong dominant component is present. Rician models also consider that the dominant wave can be a phasor sum of two or more dominant signals, e.g., the line-of-sight, plus a ground reflection. This combined signal is then mostly treated as a deterministic process and likelihood probability of having detectable signal is much more when LOS component is present.

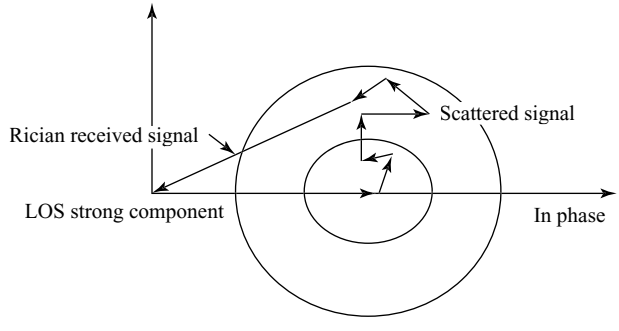


Figure 4.17 *Rician fading*

Example 4.1 For the Rayleigh fading channel with $\sigma = 0.003$ V, what would be the probability that the received voltage envelop will exceed a threshold of 6 mV and 0.1 mV? What would be the probability for the same with the presence of a direct LOS component of 6 mV?

Solution The probability of the envelope as exceeding 6 mV is given by

$$P(r \geq .006) = \int_{.006}^{\infty} \frac{r}{\sigma^2} e^{-\frac{r^2}{2o^2}} dr = 0.1353$$

$$P(r \geq .0001\text{i}) = \int_{.0001}^{\infty} \frac{r}{\sigma^2} e^{-\frac{r^2}{2o^2}} dr = 0.9994$$

$$P(r \geq .006) = \int_{.006}^{\infty} \frac{r}{\sigma^2} e^{-\frac{(r^2+A^2)}{2o^2}} I_0\left(\frac{r\,A}{\sigma^2}\right) dr = 0.6035$$

When the threshold voltage is close to 0 volt, the probability reaches to 1, as in that case it is the area under the entire Rayleigh distribution curve.

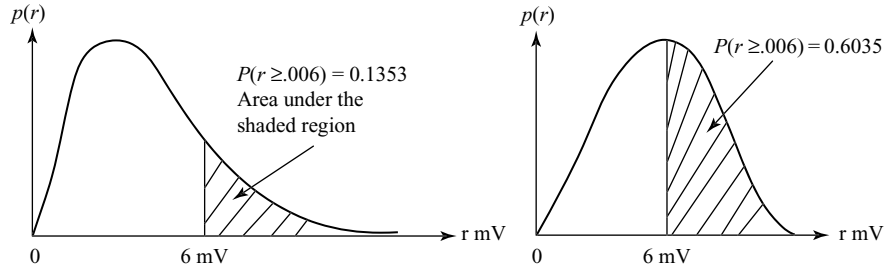


Fig. 4.18 *Rayleigh and Rician probability*

The corresponding probability of received signal voltage with the direct LOS component will be obtained from Rician distribution curve and probability for first case with LOS = 6 mV is increased to 0.6035, which justifies the theory. But for the second case probability will be same and is close to 1, as the area under the entire curve for both Rayleigh and Rician will be 1.

Example 4.2 Give the probability power distribution curve for Rayleigh channel. Find the outage probability of a Rayleigh channel if the average power is equal to 3 μW and the minimum threshold power is only 1 μW. Find the outage probabilities for P_{th} = 2 μW and 3 μW and comment on the result.

Solution The probability power distribution of Rayleigh channel is exponential as shown in the Fig. 4.19.

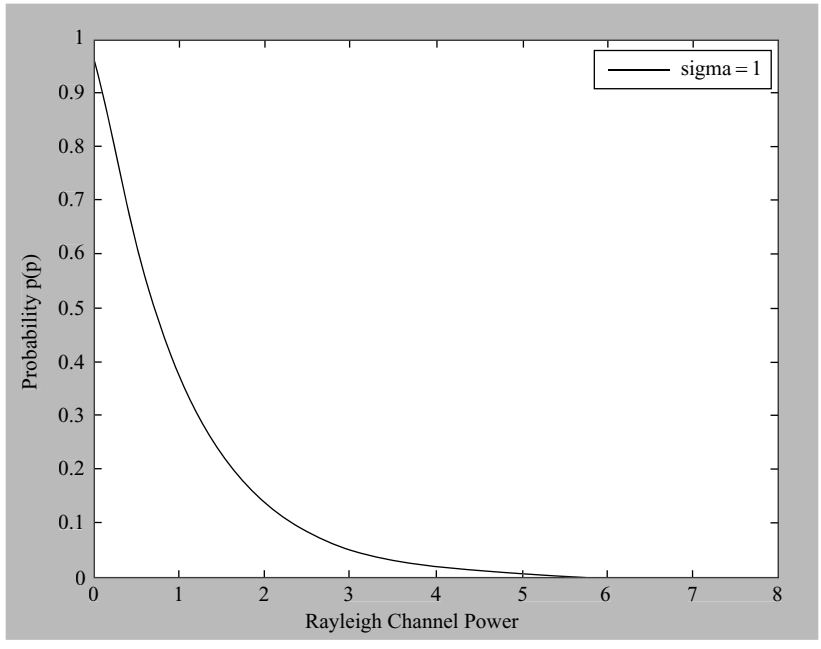


Fig. 4.19 *Probability-power distribution for Rayleigh channel with σ =1*

The outage probability is given as

$$P(p \le P_{th}) = \int_0^{P_{th}} (1/\gamma_0) e^{-\gamma/\gamma} 0 \, d\gamma,$$

where $\gamma_0 = 2\sigma^2$ is the average power for the channel, for P_{th} = 1 μW the outage probability is 0.2832. Outage probabilities for P_{th} = 2 μW and 3 μW are 0.4861 and 0.6315 respectively.

The outage probability is the area under the probability-power distribution curve with the limit from 0 to P_{th}. As threshold power increases outage probability also increases. If the received power falls below the threshold, the receiver goes into "outage" and with increased threshold chances of outage will be more.

Example 4.3 Using MATLAB plot the envelopes of received signal in a mobile environment where the carrier frequency is 2 GHz, the vehicle velocity is 50 mph, and 60 mph, the phase angles α_n and φ_n are uniformly distributed, the coefficient a_n has a Gaussian distribution with zero mean and standard deviation of $\sigma = .001$. Let varying number of scatterer be N,

Solution First converting the velocity to meters/second, we get 50 mph = 22.35 m/s. Thus, the maximum Doppler shift is

$$f_d = \frac{2\times10^9 \times 22.35}{3\times10^8} = 149 \text{ Hz}$$

Similarly, for velocity 60 mph the Doppler frequency shift is 179 Hz.

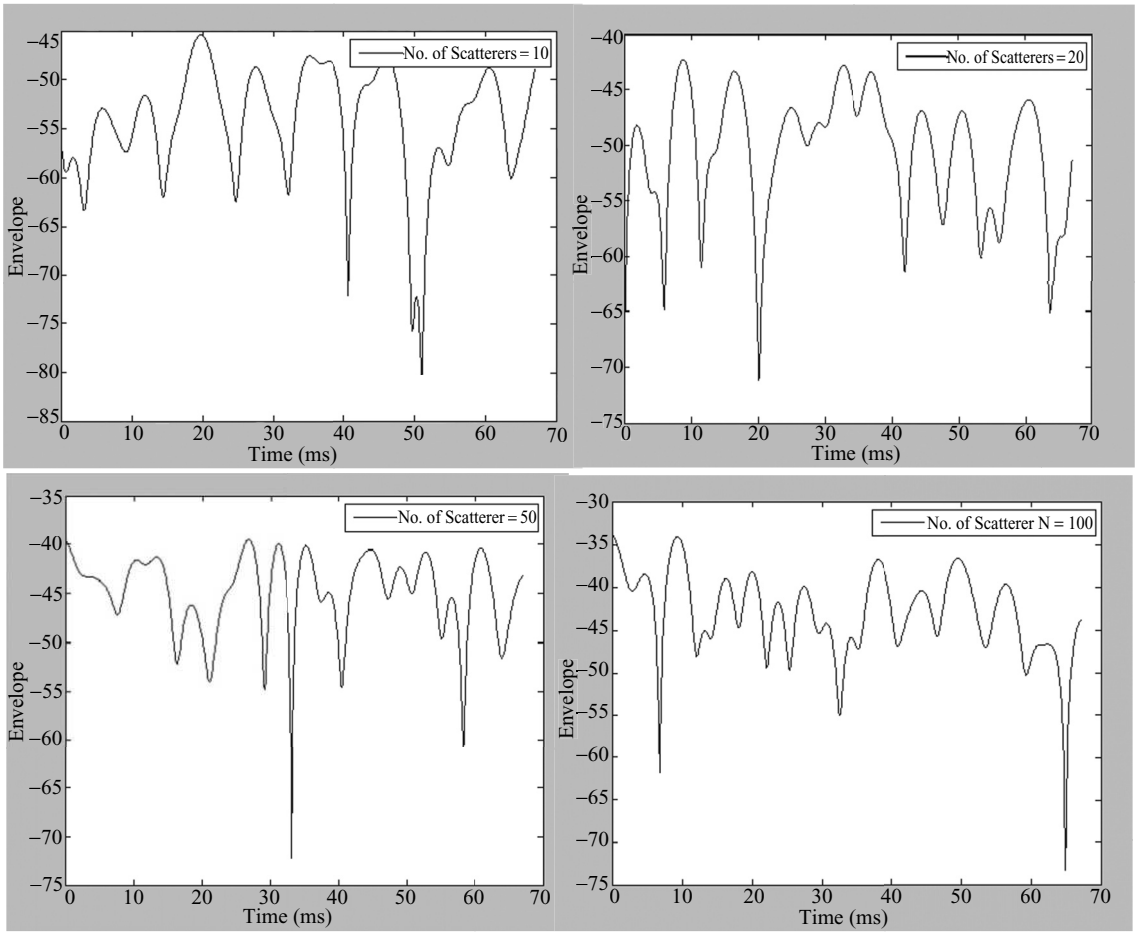


Fig. 4.20 *Envelop of the received signal 'r' in dB in Doppler fading channel, $f_d = 149$ Hz*

Using MATLAB program, we can plot the results with varied number of scatterer as given in Fig. 4.20.

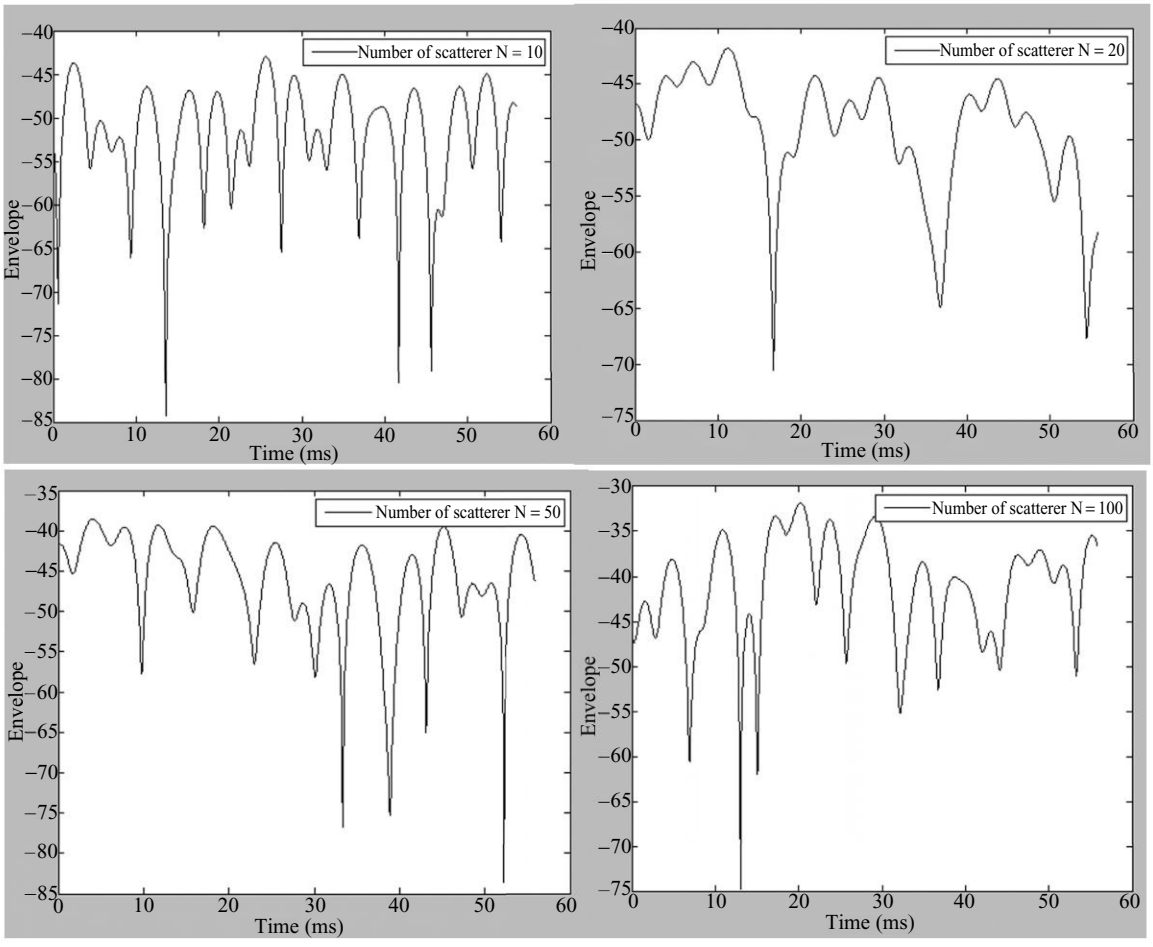


Fig. 4.21 *Envelop of the received signal 'r' in dB in Doppler fading channel, fd = 179 Hz*

In the above example in Doppler fading channel, assumptions are not realistic. In the example, the scattering from objects is angularly dependent. Thus, the coefficients a_n will be a function of time. Additionally, the phase angles α_n and φ_n change with time.

If we assume a large number of paths, the uniform distribution of angles α_n results in a sinusoidal variation in the Doppler frequencies f_n. This transformation of the random variable results in a Doppler power spectrum derived by Gans [3] and Jakes [4].

$$S_d(f) = \frac{\sigma^2}{\pi fd\sqrt{1-\left(\frac{f}{fd}\right)}} \quad |f| \le f_d \tag{4.28}$$

where $\sigma^2 = \sum_{n=1}^{N} E[a_n^2]$ = Average power signal

A plot of the Doppler power spectrum is shown in Fig. 4.22.

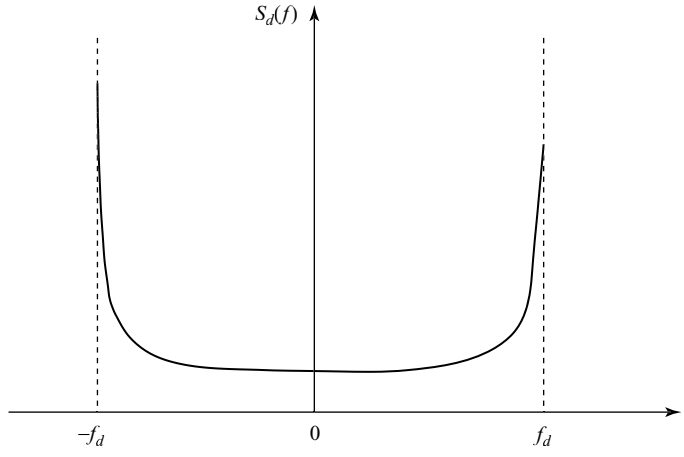


Fig. 4.22 *Doppler power density spectra*

Example 4.4 Using MATLAB, plot the Doppler Power Spectrum Density for the same problem given in (4.3) with Doppler frequency shift f_d = 149 Hz, for N =10, and N =100 to validate Eq. (4.28).

Solution In Fig. 4.20, the envelop of the received signal due to Doppler effect is plotted for different number of scatterer which shows very random nature. Taking this envelop and using Fast Fourier Transform (FFT) frequency domain representation say, Y(f) of the received signal can be obtained. Then by applying standard definition of the power spectral density = $|Y(f)|^2$ per unit bandwidth will give the PSD plot for the required problem. Figure 4.23 shows this plot which shows the exact similarity with the theoretical spectrum given in Fig. 4.22.

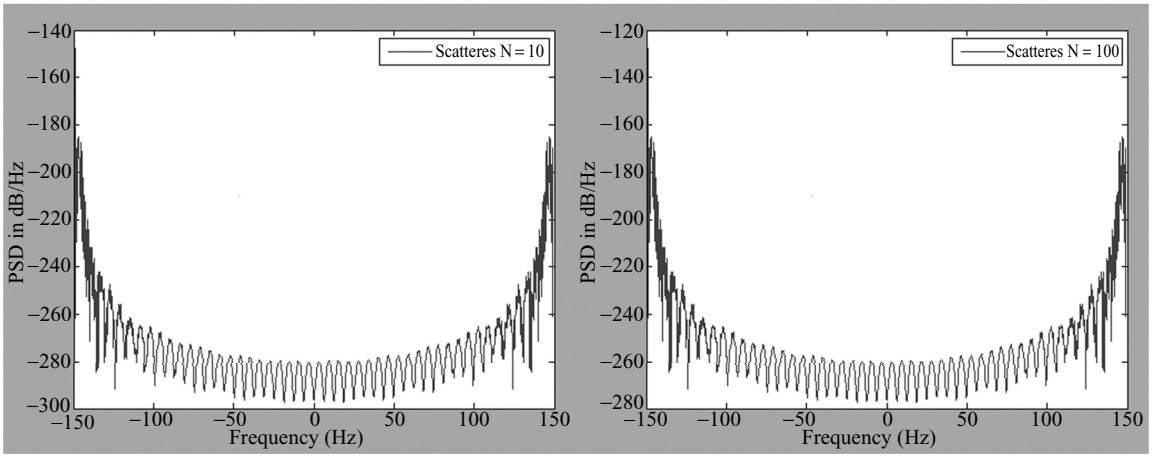


Fig. 4.23 *MATLAB plot for Doppler power spectrum density for fd =149 Hz with N = 10 and N = 100*

4.4.4 Characterization of Time Variant Channel in Terms of Correlation Functions

Bello [P.A. Bello, IEEE Transaction on Communication System, 1963] proposed a simple fading model in the year 1963 by considering the channel as ***Wide-Sense Stationary Uncorrelated Scattering (WSSUS).*** The model considers the signal variations arriving with different delays as uncorrelated. Such channel is WSS in both frequency and time domain.

When the channel changes with time randomly, the four channel characteristics $h(t,\tau)$, $H(f,t)$, $H(f,\gamma)$ and $H(\tau,\gamma)$ are random processes those vary with time and is difficult to characterize. With the assumption that RP has zero mean, then the correlation functions of the RP can be easily found with the following assumptions:

1. The channel impulse response $h(t, \tau)$ is WSS process and
2. The channel impulse responses at two different time τ_1 and τ_2 are uncorrelated if $\tau_1 \neq \tau_2$ for any t.

As long as the number of multipath components is large, using central limit theorem, it can be considered that the channel impulse response $h(t, \tau)$ is a complex Gaussian process. So, its statistical characterization is fully known from the mean, autocorrelation and cross-correlation of its in phase and quadrature components. The auto correlation function of $h(t, \tau)$ is defined as

$$\Phi_h(\tau_1, \tau_2; t, \Delta t) = E[h^*(\tau_1; t)\, h(\tau_2; t + \Delta t) \tag{4.29}$$

For WSS channel, the joint statistics of the channel is dependent only on time difference Δt. Also, for WSSUS channel, channel response due to multipath delay τ_1 is uncorrelated to channel response due to delay τ_2 if $\tau_1 \neq \tau_2$. So the autocorrelation function can be redefined as

$$\Phi_h(\tau_1; \Delta t)\delta(\tau_1 - \tau_2) \overset{\Delta}{=} \Phi_h(t; \Delta t) = E[(h^*(\tau_1; t)h(\tau_2; t + \Delta t)$$

where $\Phi_h(\tau;\Delta t)$ gives the average output power associated with the channel as a function of multipath delay $\tau = \tau_1 = \tau_2$ and the difference Δt is the observation time. This function assumes that $|\tau_1 - \tau_2| > 1/B_c$, otherwise the receiver can not resolve the two components.

The scattering function of the random channel is defined as the Fourier transform of $\Phi_h(\tau;\Delta t)$ with respect to Δt parameter and is given as

$$S_c(\tau, \gamma) = \int_{-\infty}^{\infty} \Phi_h(t; \Delta t) e^{-j2\pi\gamma\Delta t} d\,\Delta t, \tag{4.30}$$

which is the average output power associated with the channel with respect to multipath delay τ, and Doppler frequency spread γ.

The most important characteristics of the wideband channel are Power Delay Profile, Coherence bandwidth, Doppler Power Spectrum and Coherence time those are derived from the channel autocorrelation function $\Phi_h(\tau_1; \Delta t)$ and scattering parameter $S_c(\tau,\gamma)$.

Power-Delay Profile The **Power Delay Profile (PDP)** also called the **Multipath Intensity Profile** (MIP) is defined as the auto-correlation function with $\Delta t = 0$, and is denoted as $S(t) = \Phi_h(\tau; 0)$. PDP represents the average power associated with a given multipath delay. Knowledge of $S(\tau)$ helps to answer the question for a transmitted impulse, "How does the average received power vary as a function of time delay, τ?" Here the term "time delay" is used to refer to the excess delay. It represents the signal's propagation delay that exceeds the delay of the first signal arrival at the receiver. Figure 4.24 shows a typical multipath power-delay profile $S(\tau)$ versus time delay τ, assuming the minimum delay is equal to zero. For a typical wireless RF channel, the received signal usually consists of several discrete multipath components, sometimes referred as fingers. For a single transmitted impulse, the time T_m between the first and last received component represents the maximum excess delay during which the multipath signal power falls to some

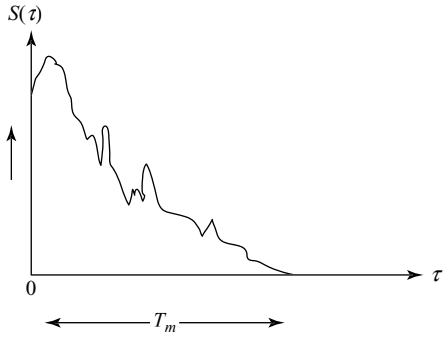


Fig. 4.24 *Power delay profile curve in a multipath channel*

threshold level below that of the strongest component. The threshold level might be chosen at 10 or 20 dB below the level of the strongest component.

The average delay $\bar{\tau}$ is typically defined in terms of PDP as

$$\bar{\tau} = \frac{\int_0^\infty \tau\, S(\tau)\, dt}{\int_0^\infty S(\tau)\, dt} = \frac{\sum_{n=1}^{N} S_n \tau_n}{\sum_{n=1}^{N} S_n} \tag{4.31}$$

As different channels with the same value of T_m can perform differently on delay power profile over the time span, T_m is not necessarily the best indicator of how any given system will perform on a channel. A more useful measurement of delay spread is most often characterized in terms of the **root mean squared delay** spread or rms delay spread that can be expressed as different ways

$$\sigma_\tau = \sqrt{\frac{\int_0^\infty (\tau - \bar{\tau})^2 S(\tau)\, d\tau}{\int_0^\infty S(\tau)\, d\tau}} = \sqrt{\frac{\sum_{n=1}^{N} S_n \tau_n^2}{\sum_{n=1}^{N} S_n} - \bar{\tau}^2}$$

$$= \sqrt{(\tau)^2 - \bar{\tau}^2} \tag{4.32}$$

where $\bar{\tau}$ is the mean delay and $(\bar{\tau})^2$ is the second order moment and is the mean squared delay, σ_τ is the square root of the second central moment of $S(\tau)$.

RMS Delay Spread: Some Typical Values

Residential building (indoor) – 5-10 ns
Office Environment – 1- 100 ns
Microcells (<500 m) – LOS 5 -100 ns, NLOS 100-500 ns
Macro Cell Urban , Suburban 100-800 ns
Bad urban, Hilly Terrain, Lossy Media > 10 μs

A channel is said to exhibit frequency-selective fading if $T_m > T_s$, where T_s is the symbol duration. This condition occurs whenever the received multipath components of a symbol extend beyond the symbol's time duration causing inter-symbol interference (ISI). A channel is said to exhibit frequency non-selective or flat fading if $T_m < T_s$. In this case, all the received multipath components of a symbol arrive within the symbol time duration.

Example 4.5 The discrete power profile delay for multipath transmission is shown in figure below, show the multipath power gain, mean delay and rms delay spread.

Solution First convert the power in terms of linear scale,
$S_1 = 10^{-(10/10)} = 0.1$ w
$S_2 = 0.316$ w, $S_3 = 0.063$ w, $S_4 = 0.0316$ w, $S_5 = 0.02$ w
The multipath power gain is S_T = Sum of all $S_n = 0.5306 = -2.75$ dB
The mean excess delay is

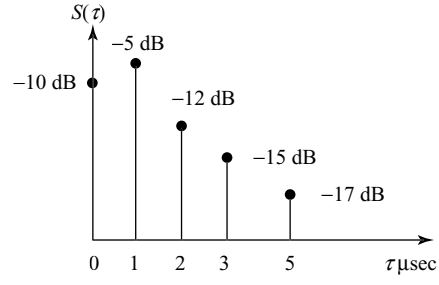


Fig. 4.25 *Discrete PDP*

$$\bar{\tau} = \frac{\sum_{n=1}^{5} S_n \tau_n}{S_T} = \left(0.1\text{x}0 + 0.316\text{x}1 + 0.063\text{x}2 + 0.0316\text{x}3 + 0.02\text{x}5\right) / 0.5306 = 1.2\ \mu\text{s}$$

The rms delay spread is

$$= \sqrt{\frac{\sum_{n=1}^{N} S_n \tau_n^2}{\sum_{n=1}^{N} S_n} - \tau^{-2}} = \sqrt{(0.316\text{x}1^2 + 0.063\text{x}\ 2^2 + 0.0316\text{x}\ 3^2 + 0.02\ \text{x}5^2)\ /0.5306 - 1.2^2}$$

$$= 1.053\ \mu\text{s}$$

Prediction of Power Delay Profile The knowledge of PDP is very much helpful in understanding the channel behavior and performance measure for equalizers and bit error rate (BER). By extensive measurements of PDP in both indoor and outdoor environment, useful models have been obtained out of which two most useful models are given here. The received power S_T is given below.

One-Sided Exponential Profile This profile is mostly described both the indoor and urban channels.

$$S(\tau) = (S_T / \sigma_\tau) e^{-\tau/\sigma} \tau \quad \tau \geq 0 \tag{4.33}$$

Gaussian Profile

$$S(\tau) = (S_T / \sqrt{2\pi}\ \sigma_\tau) e^{-\frac{1}{2}(\tau/\sigma\tau)^2} \tag{4.34}$$

Coherence Bandwidth The time varying channel can also be characterized by taking the Fourier transform of channel impulse response $h(t, \tau)$ with respect to τ, and is denoted by,

$$H(f;t) = \int_{-\infty}^{\infty} h(t,\tau) e^{-j2\pi ft} d\tau,$$

As $h(t, \tau)$ is a complex zero mean Gaussian random process in t, $H(f; t)$ is also the complex zero mean Gaussian process characterized completely by it's autocorrelation function (ACF). Thus in frequency domain for $H(f; t)$, ACF is defined as,

$$\Phi_H(f_1, f_2; \Delta t) = E[H^*(f_1;t) H(f_2; t+\Delta t) \tag{4.35}$$

$$= E[\int_{-\infty}^{\infty} h^*(\tau_1;t) e^{j2\pi f 1\tau 1}\ d\tau_1 \int_{-\infty}^{\infty} h(\tau_2; t+\Delta t) e^{-j2\pi f}\ 2^{\tau} 2\ d\ \tau_2]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E[h^*(\tau_1;t) h(\tau_2; t+\Delta t)] e^{j2\pi f 1\tau 1} e^{-j2\pi f}\ 2^{\tau} 2\ d\tau_1\ d\tau_2$$

$$= \int_{-\infty}^{\infty} \varphi_h(\tau;\Delta t)] e^{-j2\pi(f_2 - f_1)}\ d\tau = \varphi_H(\Delta f; \Delta t)$$

where $\Delta f = f_2 - f_1$, and the equality follows from the WSS and uncorrelated scattering properties of $h(\tau;t)$. In practice, $\Phi_H(\Delta f\,; \Delta t)$ is measured by transmitting two sinusoid signals separated by frequency difference

Δf and calculating the cross correlation at the receiver for the time separation Δt. Again for $\Delta t = 0$, the correlation function can be defined as,

$$\Phi_H(\Delta f; 0) = \Phi_H(\Delta f) = \int_{-\infty}^{\infty} \Phi_h(\tau) e^{-j2\pi \Delta f \tau} d\tau \tag{4.36}$$

Thus $\Phi_H(\Delta f)$ is simply the Fourier transform of power delay profile. Knowledge of $\Phi_H(\Delta f)$ helps to answer the question "What is the correlation between two received signals that are spaced in frequency Δf? $\Phi_H(\Delta f)$ also sometimes called as the spaced-frequency correlation function.

The **coherence bandwidth B_c** is defined as the frequency separation for which $\Phi_H(\Delta f) \approx 0$ for all $\Delta f > B_c$. By FT relationship between the $\Phi_h(\tau)$ and $\Phi_H(\Delta f)$, if $\Phi_h(\tau) = 0$ for some $\tau > T$ then also $\Phi_H(\Delta f) \approx 0$ for $\Delta f > 1/T$. Thus the minimum frequency separation B_c for which the channel response roughly becomes independent is the inverse of maximum time delay spread T_m of power delay profile. A more general relation with rms delay spread σ_τ is $B_c = k/\sigma_\tau$, where k depends on the shape of PDP and precise specification of B_c.

An exact relationship between coherence bandwidth and delay spread does not exist, and must be derived from signal analysis (usually using Fourier techniques) of actual signal dispersion measurements in particular channels. Several approximate relationships have been described. If coherence bandwidth is defined as the frequency interval over which the channel's complex frequency transfer function has a correlation of at least 0.9, the coherence bandwidth is approximately given by $B_c = 1/50\sigma_\tau$. A more popular approximation of B_c corresponding to a bandwidth interval having a correlation of at least 0.5 is $B_c = 1/5\sigma_\tau$.

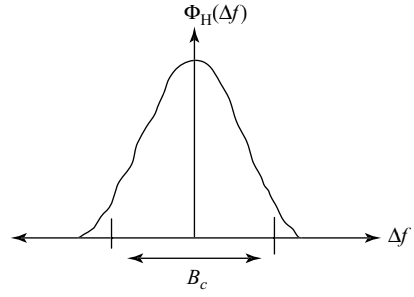


Fig. 4.26 *Distribution of spaced-frequency function and coherence bandwidth*

Equation 4.26 represents the distribution of $\Phi_H(\Delta f)$ which is the FT pair of $\Phi_h(\tau;0)$ or $S(\tau)$. The relationship of coherence bandwidth B_c is also defined in Fig. 4.26.

The coherence bandwidth B_c is a statistical measure of the range of frequencies over which the channel passes all spectral components with approximately equal gain and linear phase. Thus, the coherence bandwidth represents a frequency ranges over which frequency components have a strong potential for amplitude correlation. That is signal's spectral components over the range is affected in a similar manner, as for example, exhibiting fading or no fading. The coherence bandwidth is the reciprocal of the maximum delay spread T_m in a multipath time varying channel.

Doppler Power Spectrum and Channel Coherence Time The motion of the transmitter or receiver causes the Doppler shift in the received signal. The characterization of the LTV channel with respect to Doppler effect can be obtained by taking the Fourier transform of $\Phi_H(\Delta f; \Delta t)$ with respect to time variation Δt. Let the new function is defined as

$$S_H(\Delta f; \gamma) = \int_{-\infty}^{\infty} \Phi_H(\Delta f; \Delta t) e^{-j2\pi\gamma \Delta t} d\Delta t \tag{4.37}$$

For a single frequency Doppler effect, $\Delta f = 0$, and $S_H(\Delta f;\ \gamma) = S_H(0\ ;\ \gamma)$

$$S_H(\gamma) = \int_{-\infty}^{\infty} \Phi_H(\Delta t) e^{-j2\pi\gamma \Delta t} d\Delta t$$

$\Phi_H(\Delta f; \Delta t) = \Phi_H(\Delta t)$ is the auto-correlation function defining how the channel impulse response decorrelates over time. $\Phi_H(\Delta t = T) = 0$, indicates the channel impulse response at times separated by T are uncorrelated and

therefore independent as the channel is Gaussian random process. Thus the channel coherence time T_c is the time range over which $\Phi_H(\Delta t)$ is approximately non zero. Time varying channel de-correlates after time T_c. The function $S_H(\gamma)$ is called the Doppler power spectrum of the channel. The maximum γ value for which $|S_H(\gamma)|$ is zero is known as the Doppler frequency spread and is defined by f_d and it is the inverse of the coherence time T_c. Figure 4.27 is the representation of $S_H(\gamma)$ with respect to Doppler frequency change γ. The function $\Phi_H(\Delta t)$ is the autocorrelation function of a channel's response to a sinusoid and is known as the time-spaced function.

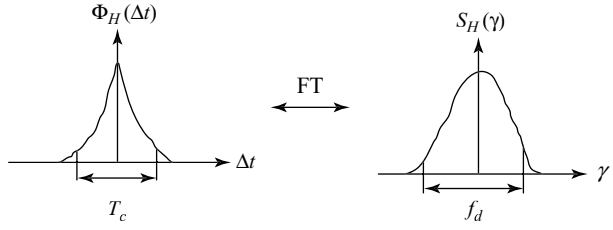


Fig. 4.27 *Doppler power spectrum, Doppler spread and coherence time*

The relationship between the scattering parameters and the four different channel response is given in Fig. 4.28.

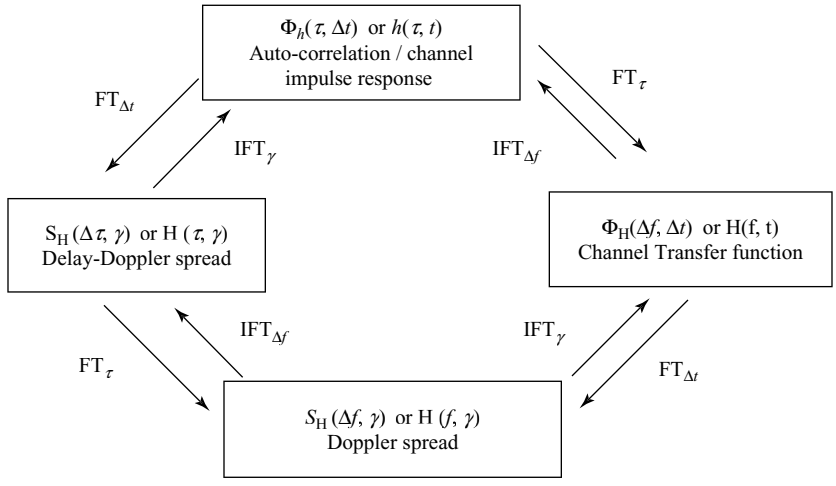


Fig. 4.28 *Relationship between the channel functions and Fourier transform*

Coherence Bandwidth: Defines a frequency range where the correlation coefficient is greater than a given threshold

$B_c \geq 1/2\pi\sigma_\tau$, where σ_τ is the rms delay spread

Coherence Time: The time duration over which two received signals has an amplitude correlation greater than a given threshold. It is the statistical measure of the time duration over which the channel impulse response is essentially invariant.

$T_c \approx 1/2\pi f_d$, where f_d is Doppler frequency shift

Example 4.6 Consider a wideband channel with multipath delay profile as given below.

$$S(\tau) = \begin{cases} e^{-\tau/.00001} & 0 \le \tau \le 10\ \mu\text{sec}. \\ 0 & \text{else} \end{cases}$$

Find the mean delay, rms delay spread and the maximum symbol rate that a linear modulated signal can be transmitted without ISI.

Solution The mean delay is given by

$$\bar{\tau} = \frac{\int_0^\infty \tau\, S(\tau)\, d\tau}{\int_0^\infty S(\tau)\, d\tau} = 4.18\ \mu\text{s}$$

Rms delay is given by

$$\sigma_\tau = \sqrt{\frac{\int_0^\infty (\tau - \bar{\tau})^2 S(\tau)\, d\tau}{\int_0^\infty S(\tau)\, d\tau}} = 2.817\ \mu\text{s}$$

To avoid ISI, the symbol rate T_s has to be greater than rms delay spread. Let $T_s > 10\ \sigma_\tau$ which is equal to $1/10\ \sigma_\tau$. So the symbol rate is $1/T_s = 35.5$ kbps.

Space–Time Channel Model—The Power Angular Profile The PDP along with the estimation of different path delays are helpful to characterized the channel and to calculate the channel bandwidth. The PDP is especially relevant for *single-input single-output* (SISO) channels since one can view the impulse response as being for a SISO system. In today's wireless communication system with higher transmission rate suffers of ISI effects. To overcome the bad effect of ISI, diversity antenna system is used. As, for example, MIMO (multiple input multiple output) system is the most common use to broadband wireless access networks. If an array is used at the receiver end then the gain of the array is dependent on the angle of arrival (AOA). System with multiple antennas requires channel characterization in terms of spatial (AOA) as well as temporal behavior. So, it is instructive to have the knowledge of angular spread and mean angle of arrival. So the angular equivalent of the PDP is defined with PAP, Power Angular Profile which is given as

$$S(\theta) = \sum_{n=1}^{N} S_n\ \delta(\theta - \theta_n) \tag{4.38}$$

where θ is the AOA of the multipath relative to the origin of the antenna array. $S(\theta)$ is the average received signal power as function of θ. The concept of a PAP conveys angular impulse response information, which assists in channel characterization. The mean angular spread and rms delay spread are expressed as given below.

Mean of angular spread,

$$\bar{\theta} = \frac{\int_{-\pi}^{\pi} \theta\, S(\theta)\, d\theta}{\int_{-\pi}^{\pi} S(\theta)\, d\theta} = \frac{\sum_{n=1}^{N} S_n\ \theta_n}{\sum_{n=1}^{N} S_n} \tag{4.39}$$

The rms angular spread,

$$\sigma_\theta = \sqrt{\frac{\int_{-\pi}^{\pi}(\theta-\bar{\theta})^2 S(\theta)\,d\theta}{\int_{-\pi}^{\pi} S(\theta)\,d\theta}} = \sqrt{\frac{\sum_{n=1}^{N} S_n\,\theta_n^2}{\sum_{n=1}^{N} S_n} - \bar{\theta}^2}$$

$$= \sqrt{\langle\theta^2\rangle - \bar{\theta}^2} \tag{4.40}$$

The two signals received at AOAs separated by $1/\sigma_\theta$ is uncorrelated.

Example 4.7 Given that the coherence BW, $B_C = \frac{1}{5\sigma_\tau}$, where σ_τ is the rms delay spread. Show that a flat fading channel occurs when, $T_S \geq 10\sigma_\tau$, T_s is the reciprocal of the baseband signal BW.

Solution The coherence BW, $B_c = \frac{1}{5\sigma_\tau}$ where σ_τ is the rms delay spread. Here, Ts is the reciprocal of the baseband signal BW.
In this case, we assume that the given signal is a RF signal in which we assume null-to-null bandwidth.
The condition for flat fading that is required to be satisfied is

$$B_c \approx \frac{1}{5\sigma_\tau} \geq 2\ Baseband \cong \frac{2}{T_s}$$

Therefore,

$$\frac{2}{T_s} \leq \frac{1}{5\sigma_\tau}$$

or, $T_s \leq 10\sigma_\tau$ for flat fading (Proved).

Example 4.8 Find the mean and rms delay spread $\bar{\tau}$ and σ_τ for multipath propagation for which power delay profile is given. Now if a particular modulation provides suitable BER performance whenever, $\sigma_\tau / T_s \leq \frac{1}{10}$, determine the smallest symbol period T_S (and thus the greatest, $R_S = 1/T_S$ symbol rate) that may be sent through RF channels without using equalizer.

Solution The mean excess delay for a multipath propagation is given as

$$\bar{\tau} = \frac{\sum_k P(\tau_k)\tau_k}{\sum_k P(\tau_k)}$$

The RMS delay spread $\sigma_\tau = \sqrt{\overline{\tau^2} - (\overline{\tau})^2}$

$$\overline{\tau^2} = \frac{\sum_k P(\tau_k)\tau_k{}^2}{\sum_k P(\tau_k)}$$

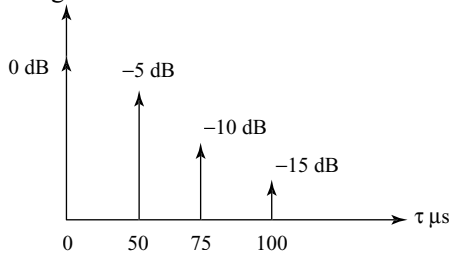


Fig. 4.29 *Power delay profile for the problem 4.8*

Hence,

$$\bar{\tau} = \frac{.316\times 50 + .1\times 75 + .0316\times 100}{1 + .316 + .1 + .0316} = 18.276\ \mu s$$

Similarly $\bar{\tau}^{-2} = \dfrac{.316\times 50^2 + .1\times 75^2 + .0316\times 100^2}{1 + .316 + .1 + .316} = 1152.5974$

Considering again, $\sigma_\tau = \sqrt{\bar{\tau}^{-2} - (\bar{\tau})^2} = 28.611\ \mu s$

We have, $\sigma_\tau \le \dfrac{1}{10} T_s$

So, $T_s \ge 10\sigma_\tau$, or $T_s = 286.11\ n s$

Hence, $R_{s\max} = \dfrac{10^9}{286.11} = 3.495\ Mbps.$

Example 4.9 Given the Doppler frequency spread B_d, $f_m = \dfrac{Vf_c}{c}$, $v \rightarrow$ velocity of mobile, $f_c \rightarrow$ carrier frequency, c$\rightarrow$ velocity of light. The coherent time $T_C = \dfrac{9}{16\pi f_m}$

Now if a baseband binary message with bit rate R_b = 100 Kbps is modulated by an RF carrier using BPSK, answer the following:

(a) Find the range of values required for the RMS delay spread of the channel such that the received signal is flat fading signal
(b) If f_c = 5.8 GHz, what is T_C, speed of vehicle 30 miles/hour.
(c) For your answer in (b), is the channel "fast" or "slow" fading?
(d) Given your answer in (b), how many bits are sent while the channel appears "static"?

Solution $R_b = 100 Kbps$ (*given*)

(a) So, $T_s = \dfrac{1}{100000} S = 10^{-5} S$

We already know the condition for flat fading to be $\sigma_\tau \le \dfrac{1}{10} T_S$.

Hence, $T_S \ge 10\sigma_\tau$ or, $\sigma_\tau \le \dfrac{1}{10} T_S$.

Hence, the required range of value is $0 \le \sigma_\tau \le 10^{-6} S$.

(a) It is given that f_c = 5.8 GHz.

Speed of vehicle = 30 miles/hr $= \dfrac{30\times 1\times 5280\times 12\times 2.54}{3600\times 100\times 1\times 1\times 1} \dfrac{m}{s} = 13.4 \dfrac{m}{s}$

$$So,\ f_m = \frac{Vf_c}{c} = \frac{13.4\times 5.8\times 10^9}{3\times 10^8} = 259.28 = 259\ \text{Hz} \cong 260\ \text{Hz}$$

$$T_C = \frac{9}{16\pi f_m} = \frac{9}{16 \times 3.14 \times 2.60} \text{s} = 6.89 \times 10^{-4} s$$

(b) Here we find that $T_S < T_c$, hence the channel has 'slow fading'.

(c) Number of bits sent $= R_b.T_C = \frac{10^5 b}{s} \times 6.89 \times 10^{-4} \cong 68.9 \cong 69.$

4.5 PROPAGATION MODELS FOR WIRELESS NETWORKS

Land-mobile communication is controlled with particular propagation complications compared to the channel characteristics in radio systems with fixed and carefully positioned antennas. The antenna height at a mobile terminal is usually very small, typically less than a few meters. Hence, the antenna is expected to have very little clearance, so obstacles and reflecting surfaces in the proximity of the antenna have a potential influence on the characteristics of the propagation path. Moreover, the propagation characteristics change from place to place and, from time to time as the mobile unit moves. Thus, the transmission path between the transmitter and the receiver can vary from simple direct line of sight to one that is severely obstructed by buildings, foliage and the terrain placing a limit for the mobile radio channel on the performance of wireless systems. The wireless media or propagation channel consists of different objects, large building, and trees, etc., that scatter the radio-transmitted signal. Scatterer provides various channel impairments including fading, multipath delay spread, Doppler spread, attenuation etc in addition to background noise. Though the well-known Maxwell differential equations provide solution to all exhaustive electromagnetic problems and thus in radio propagation, unfortunately due to massive complexity of these solution they are only of practical use in certain simple and idealized situations. Radio engineers tend to use simpler models for wireless propagation physical phenomena when viewed from a system perspective.

The mobile radio channel is usually evaluated from statistical propagation models, no specific terrain data is considered, and channel parameters are modeled as stochastic variables. The mean signal strength for an arbitrary transmitter-receiver (Tx-Rx) separation is useful in estimating the radio coverage of a given transmitter whereas measures of signal variability are key determinants in system design issues such as antenna diversity and signal coding.

Three mutually independent, multiplicative propagation phenomena can usually be distinguished -multipath fading, shadowing, and large-scale path loss.

If the Tx-Rx separation (several hundreds of meters) is useful in determining the mean signal strength, then the propagation model is known as **large-scale model**. In contrary, when propagation models that are characterized by the rapid fluctuation of the received signal strength over the short travel distance (of the order of few wavelengths) due to the movement of mobile terminal or over a small time duration of the order of seconds, are known as **small-scale fading**. In a small scale fading, when mobile terminal moves over a fraction of wavelength, the received signal power may vary 3 to 4 order of magnitudes. This is due to the fact that received signal power is the sum of the many contribution coming from different directions and the phase of the received signal are varied due to different propagation paths. For large-scale propagation, local average received signal is computed by averaging signals over 5λ to 40λ.

The large-scale effects of path losses cause the received power to vary gradually due to signal attenuation determined by the geometry of the path profile in its entirety. This is in contrast to the local propagation mechanisms, which are determined by building and terrain features in the immediate vicinity of the antennas.

The large-scale effects determine a power level averaged over an area of tens or hundreds of meters and therefore called the area-mean power. Shadowing introduces additional fluctuations, so the received local-mean power varies around the area mean. The term local mean is used to denote the signal level averaged over a few tens of wavelengths, typically 40 wavelengths. This ensures that the rapid fluctuations of the instantaneous received power due to multipath effects are largely removed.

Multipath leads to rapid fluctuations of the phase and amplitude of the signal if the vehicle moves over a distance in the order of a wavelength or more, multipath fading thus has a small-scale effect.

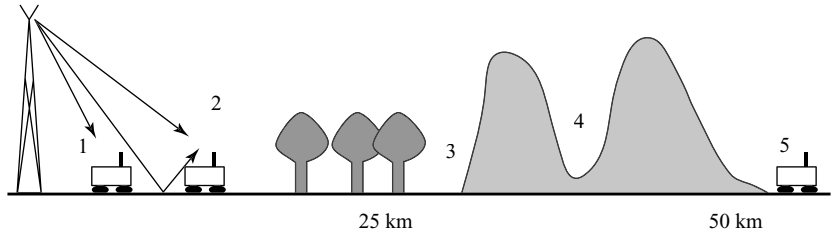


Fig. 4.30 *Location dependent propagation path loss model*

The most appropriate path loss model depends on the location of the receiving antenna. With respect to Fig. 4.30, the following divisions can be made:

1. Location 1 corresponds to free space loss for an accurate estimate of path loss.
2. Location 2, a strong line-of-sight is present, but ground reflections can significantly influence path loss. The two-ray path loss model is applicable for this case
3. Location 3, significant diffraction losses are there due to the presence of long trees, so plane earth loss needs to be corrected
4. Location 4, a simple diffraction model is suitable to give an accurate estimate of path loss.
5. Location 5, loss prediction becomes difficult and unreliable since multiple-diffraction is involved within the path.

4.5.1 Free Space Propagation Model

When the transmitting and receiving antennas are placed over a large distance (compared to wavelength) and the received signal is in LOS path, unobstructed, then the propagation model is the free space model. Satellite communication and microwave LOS radio links undergo this type of propagation. This model is the basis of understanding received signal power at a separation distance r between the transmitter and receiver antennae and is governed by the Friis free space equation,

$$P_{rx}(r) = \frac{P_{tx}\, G_{tx}\, G_{rx}\, \lambda^2}{(4\pi)^2 r^2 L} \tag{4.41}$$

where P_{rx} = received power in watt, function of separation distance r
P_{tx} = Transmitted power in watt
G_{tx} and G_{rx} gain of the transmitting and receiving antenna, respectively (isotropic)
L = System loss factor not related to propagation and is greater than 1,
λ = Wavelength in metres = c/f (velocity of light in free space/ frequency of operation).

The free space model basically represents the communication range as a circle around the transmitter. If a receiver is within the circle, it receives all communications.

The surface area of a sphere of radius r is $4\pi r^2$, so that the power flow per unit area S is in watts/meter^2 at distance r from a transmitter antenna with input accepted power P_{tx} and antenna gain G_{tx} [13],

$$S = P_{tx}\, G_{tx}/(4\pi r^2) \tag{4.42}$$

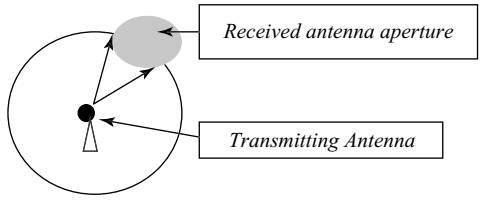


Fig. 4.31 *Free space propagation*

Transmitting antenna gain is defined as the ratio of the intensity (or power flux) radiated in some particular direction to the radiation intensity that would be obtained if the power accepted by the antenna were radiated isotropically. When the direction is not stated, the power gain is usually taken in the direction of maximum power flow. The product $G_{tx}P_{tx}$ is called the effective radiated power (ERP) of the transmitter. The available power P_{rx} at the terminals of a receiving antenna with gain G_{rx} is

$$P_{tx}(r) = \frac{P_{tx}\ G_{tx}\ A_{\text{eff}}}{(4\pi)r^2} = (\lambda/(4\pi r)^2 G_{rx} P_{tx} G_{tx} \tag{4.43}$$

where A_{eff} is the effective area or aperture of the antenna and $G_{rx} = 4\pi A_{\text{eff}}/\lambda^2$ and L = 1

Expressing in dB, i.e., 10 $\log_{10}$,

$$P_{rx}(\text{dBW}) = P_{tx}(\text{dBW}) + G_{tx}(\text{dBi}) + G_{rx}(\text{dBi}) + [20\log_{10}(\lambda/4\pi) - 20\log_{10}(r)]$$

Combining the last two terms of the above equation provides Path Loss (PL) for free space propagation. This is the channel's loss in going from the transmitter to the receiver expressed in dB. The first two right hand terms combined is called Effective Isotropic Radiated Power or EIRP. EIRP is the equivalent transmitter power required if an isotropic (0 dBi) antenna were used. Using these definitions, for free space, propagation path loss equation can be written as

$$P_{rx}(\text{dBW}) = \text{EIRP} + G_{rx}(\text{dBi}) - P_L(\text{dB}) \tag{4.44}$$

In practice, ERP (effective radiated power) is more used than EIRP to denote the maximum radiated power with respect to dipole antenna that has gain 2.64 (2.15 dB). Thus ERP is 2.15 dB smaller than EIRP for same transmission system.

In general, the path loss is defined as the 10 $\log_{10}$ (effective transmitted power/received power)

$$\text{So, } P_L = 10\log_{10}(P_{tx}/P_{rx}) \tag{4.45}$$

The Friis free space model is valid only for the received power when the distance between the antennas are in far field region. So, the received power is considered with respect to a reference distance d_0 such that $r > d_0$ lies in the far field region (>2D$^2/\lambda$).

Considering the reference distance, the received power is

$$P_{rx}(r) = P_{rx}(d_o)(d_o/r)^2 = P_0(d_o/r)^2 \qquad r > d_o > 2D^2/\lambda \tag{4.46}$$

The received power has large dynamic range and varied many orders of magnitude over the coverage area of the mobile wireless systems. So, sometimes received power can be measured in dBm or dBW corresponding to 1 mW or 1W power respectively. As for example, when P_{rx} is measured in dBm, it can be expressed as,

$$P_{rx}(r) = P_{rx}(d_0)(d_0/r)^2 = P_0(d_0/r)^2$$

where $P_0 = P_{rx}(d_0)$ =Power received at reference distance.

$$P_{rx}(r)(\text{dBm}) = 10\log_{10}[P_{rx}(d_0)/0.001\text{W} + 20\log_{10}[d_0/r], \quad r > d_0 > 2D^2/\lambda, \tag{4.47}$$

Typical values of d_0 for low gain antennas in the frequency range 1-2 GHz region are 1-100 m in indoor environment and 100 m to 1 km to outdoor environment.

Example 4.10 If the received power at a reference distance d_0 = 1 km is equal to 1 μwatt, find the received power at distances of 2 km, 4 km and 8 km from the same transmitter for the following path loss model:

(a) Free space

(b) κ =3 and 4, comment on the results

Solution

(a) We know from Eq. (4.46), for free space $P_{rx}(r) = P_{rx}(d_0)\ (d_0/r)^2 = P_0\ (d_0/r)^2$

Given $P_0 = 10^{-6}$ watt = –30 dBm, here $d_0 = 1$ km

For $d = 2$ km, $P_{rx} = 10^{-6}\ (1/2)^2 = -66$ dB = –36 dBm

(= 20 $\log_{10}$ (d_0 /r) + 10 $\log_{10}$ 10^{-6} = 20 $\log_{10}$ (1/2) + (10 x –6) = -66.02 dB)

Similarly, P_{rx} (at r = 4 km) = $10^{-6}\ (1/4)^2 = -$ 72 dB = –42 dBm

P_{rx} (at r = 8 km) = $10^{-6}\ (1/8)^2 = -$ 78 dB = –48 dBm

(b) With the path loss exponent κ = 3 and 4, the received power will vary according to κ, but not as inverse distance square law.

For κ = 3, $P_{rx} = 10^{-6}$ x $(1/r)^3$ and for κ = 4, $P_{rx} = 10^{-6}$ x $(1/r)^4$

$P_{rx}(r=2) = 10^{-6} \times (1/2)^3 = -69$ dB $\quad P_{rx}(r=2) = 10^{-6} \times (1/2)^4 = -72$ dB

$P_{rx}(r=4) = 10^{-6} \times (1/4)^3 = -78$ dB $\quad P_{rx}(r=2) = 10^{-6} \times (1/4)^4 = -84$ dB

$P_{rx}(r=8) = 10^{-6} \times (1/8)^3 = -87.09$ dB $\quad P_{rx}(r=2) = 10^{-6} \times (1/8)^4 = -96.12$ dB

As the path loss exponent increases, the received power strength for the same distance is decreased sharply resulting faster power loss.

4.5.2 Two-ray Ground Reflection Model

A single line-of-sight path between two mobile nodes is often the only means of propagation. The two-ray ground reflection model considers both the direct path and a ground reflection path. The two-ray model of the mobile channel is based on the diagram as shown in Fig. 4.32 with base station antenna height h1, mobile station antenna height h2 and with the separation distance between base station and mobile station '*d*'.

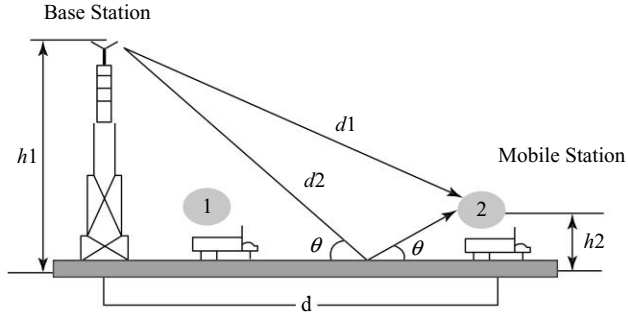


Fig. 4.32 *Two-ray model*

The distance traveled by the direct and indirect ray after the reflection from ground is given respectively by the expressions,

$$d1 = \sqrt{(h1 - h2)^2 + d^2} \quad \text{and} \quad d2 = \sqrt{(h1 + h2)^2 + d^2}$$

The difference of path between the traveled rays is given by

$$\Delta d = d2 - d1 = d[\sqrt{1 + (h1 + h2)^2/d^2} - \sqrt{1 + (h1 - h2)^2/d^2}\,]$$

This can be simplified by considering that mobile station is far from the base station and in this situation it can be approximated as

$$(h1 + h2)/d << 1 \text{ leading } \Delta d \cong 2h1h2/d$$

The received power by two-ray reflection model is given by

$$P_{r\_2\text{ray}} = P_t(\lambda/4\pi d)^2 \,|1 + Ae^{j\Delta\phi}\,|^2 \tag{4.48}$$

Where P_t transmitted power, A = -1 is the ground reflection coefficient and $\Delta\phi$ phase difference caused by the difference of path traversed.

$$\Delta\phi = \beta\,\Delta d = 2\pi\,\Delta\, d/\lambda \cong 4\pi\, h1\, h2/\lambda d$$

$$P_{r\_2\text{ray}} = P_t(\lambda/4\pi d)^2 \,|1 - e^{j\Delta\phi}\,|^2$$

Which can be simplified even further for $\Delta\phi << 1$, resulting in

$$P_{r\_2\text{ray}} = P_t(\lambda/4\pi d)^2(\Delta\phi)^2 = P_t(h1\, h2/d^2)^2 \tag{4.49}$$

The received power under this model condition is proportional to the square of the antenna heights and decreases with the fourth power of the distance between them. The above equation shows a faster power loss at a rate of 40 dB/decade as the distance increases. For large values of *d*, the received power and path loss become independent of frequency.

The path loss for two-ray model with antenna gains can be expressed in dB as

$$P_L(\text{dB}) = 40\log_{10} d - (10\log_{10} G_t + 10\log_{10} G_r + 20\log_{10} h_t + 20\log_{10} h_r) \tag{4.50}$$

G_t and G_r are the gain of transmitting and receiving antennas, h_t and h_r heights of the transmitting and receiving antennas respectively.

However, the two-ray model does not give a good result for a short distance due to the oscillation caused by the constructive and destructive combination of the two rays. Instead free space model is still used when *d* is small.

Example 4.11 In the two-ray path loss model in Fig. 4.32 if the height of the antenna is h_t and height of the receiver is h_r, then prove that $(d2 - d1) \cong 2\, h_t\, h_r\, /d$, given that $d >> h_t$ and $d >> h_r$.

Solution: With respect to Fig. 4.22, we can write

$$d2 - d1 = d\left[\sqrt{1 + (h_t + h_r)^2/d^2} - \sqrt{1 + (h_t - h_r)^2/d^2}\right]$$

With the condition $d >> h_t$ and $d >> h_r$ it can be written as

$(h_t + h_r) / d << 1$ and $(h_t - h_r) / d << 1$

From the Taylor's series expansion of a function $f(x+h)$ with $h << 1$,

$f(x+h) = f(x) + hf'(x) + h^2/2!\, f''(x) + \ldots\ldots$

Let, $f(x+h) = \left[\sqrt{1 + (h_t + h_r)^2/d^2}\right]$

Neglecting the higher order terms,

Then, $d2 - d1 \approx d\,[\,1 + (1/2)(h_t + h_r)^2/d^2 - 1 - (1/2)(h_t - h_r)^2/d^2\,]$

$\approx d$ x 4 $h_t h_r /2d^2$
$\approx 2\ h_t h_r /d$ (verified)

Example 4.12 In a two-ray ground refection model as shown in Fig. 4.33 below, assume that the $\Delta\phi$ (phase difference) must be kept below 5.9861 radians for phase cancellation. Assume that the receiver antenna height of 1.9 m and given a requirement that incident angle θ_i be less than 5^0, then

(a) Calculate minimum allowable values for the Tx-Rx separation distance and the height of the base transmitter antenna when the carrier frequency is 900 MHz.
(b) Find the time delay between the two rays at the receiver.

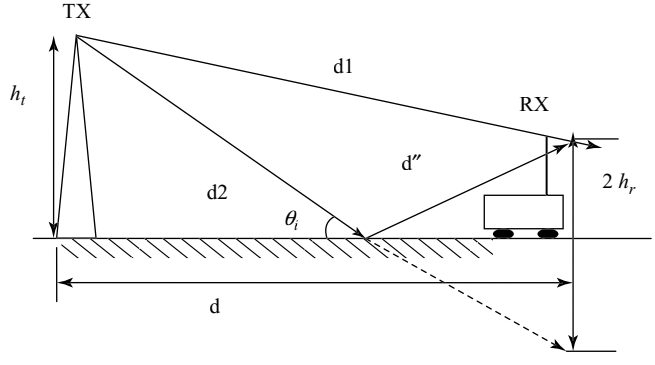


Fig. 4.33 *Two–ray model for Problem 4.12*

Solution

(a) Considering $d >> (h_t + h_r)$, $\Delta\phi = (2\pi/\lambda) \times 2\ h_t h_r /d$
$\Rightarrow d = (4\pi/\lambda) \times h_t h_r / \Delta\phi$

From the problem definition,

$\tan \theta_i = (h_t + h_r)/d < \tan 5^0$
Putting d, $\Rightarrow (h_t + h_r) / \{(4\pi/\lambda)$ x $h_t h_r / \Delta\phi\} < \tan 5^0$
$h_t > h_r / [((4\pi/\lambda)$ x $h_r \tan 5^0 / \Delta\phi) -1]$,
here $h_r = 1.9$ m, $\Delta\phi = 5.9861$, $\lambda = c/f = 0.333$ m
$h_{t\text{-min}} = 38.0$ m
$d_{\min} = (4\pi/\lambda)$ x $h_{t\text{-min}} h_r / \Delta\phi$
$d_{\min} = 454.7$ m

So, the minimum separation between the Tx-Rx is 454.7 m and height of the base transmitter is 38.0m.

(b) Time delay $t_d = \Delta\phi / 2\pi f_c = 5.9861/2$x3.14x 900x$10^6$ s = 1.0585 x 10^{-9} s.

Example 4.13 If the received power at a reference distance $d_0 = 1$ km is equal to 1 μwatt, find the received power at distances of 2 km, 4 km and 8 km from the same transmitter using exact expression for two-ray ground reflection model. Given height of transmitting antenna is $h_t = 40$ m, receiving antenna $h_r =$ 3m, $G_t = G_r = 0$ *dB*, operating frequency $f = 1800$MHz. Give a comparative table for the received power with respect to Problem 3.8 for free space model.

Solution The expression for received power from a transmitter of power P_t at a distance d_0 is

$$P_r(d_0) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d_0^2}$$

So,

$$P_t = \frac{P_r \times (4\pi)^2 d_0^2}{G_t G_r \lambda^2}$$

Given, $P_r = 10^{-6}$ Watt, $d_0 = 1$ km, $G_t = G_r = 0$ dB = 1, $\lambda = c/f = 3 \times 10^8 / 18 \times 10^8 = 0.1667$ m. Putting all the values in the expression for P_t, we get $P_t = 5.679$ kW

Using two ray ground reflection model the expression for exact received power at a distance d is

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2} \times (\Delta\phi)^2 = P_t \times (h_t\ h_r / d^2)^2$$

where $\Delta\phi = 4\pi \times h_t\ h_r / \lambda d$

$$P_r(d) = P_t \times (h_t\ h_r)^2 \times (1/d^4) = 5.679 \times 10^3 \times (40 \times 3)^2 \times (1/d^4) = 81.7776 \times 10^6 / d^4 \text{ W}$$

$$P_r(d = 2 \text{ km}) = 81.7776 \times 10^6 / (2 \times 10^3)^4 = 10^{-6} \times 81.7776/2^4$$

$$= 5.111 \times 10^{-6} \text{W} = -58.29 \text{ dB} = -28.29 \text{ dBm}$$

$$P_r(d = 4 \text{ km}) = 81.7776 \times 10^6 / (4 \times 10^3)^4 = 10^{-6} \times 81.7776/4^4$$

$$= 0.3194 \times 10^{-6} \text{W} = -64.95 \text{ dB} = -34.95 \text{ dBm}$$

$$P_r(d = 8 \text{ km}) = 81.7776 \times 10^6 / (8 \times 10^3)^4 = 10^{-6} \times 81.7776/8^4$$

$$= 0.000244 \times 10^{-6} \text{W} = -96.12 \text{ dB} = -66.12 \text{ dBm}$$

Distance in km	Received power by free space model in dBm	Received power by two ray model in dBm
2.0	-36	-28.29
4.0	-42	-34.95
8.0	-48	-66.12

4.5.3 Distance Power Loss

The distance dependence path loss is basically dependent on three components corresponding to three different granularities in the distance 'r' where the mobile terminal is presently located from the access point. The three main components are:

1. Distance dependent average path gain (G_{av})
2. Shadowing gain (G_s)
3. Multi-path gain (G_{mp})

The received power at a point is given by

$$P_R = P_t\, G_{av}\, G_s\, G_{mp} \tag{4.51}$$

In dB this can be written as

$$P_R(\text{dB}) = P_t(\text{dB}) + G_{av}(\text{dB}) + G_s(\text{dB}) + G_{mp}(dB) \tag{4.52}$$

where P_t is the transmitting power.

The distance dependent average path loss can be modeled with some simple expression like,

$$G_{av} = C/r^k \tag{4.53}$$

in dB the model is known as the Okumura-Hata model [8].

$$G_{av}(\text{dB}) = 10 \log_{10} (C/r^k) = C' - 10 \times \kappa \log_{10} r \tag{4.54}$$

The exponent κ is empirically determined by experimentation over different terrains. Theoretical κ values ranges from 2 and 4 for free space and plane, smooth, perfectly conducting terrain respectively. Typical values for irregular terrain are between 3.0 and 3.4 and in forestall terrain propagation can be appropriately described as in free space plus some diffraction losses, but without significant ground-wave losses (κ = 2.6). If the propagation model has to cover a wide range of distances, κ may vary, as different propagation mechanisms dominate at different ranges. In micro-cellular networks, κ typically changes from approximately 2 to approximately 4 at some turnover distance 'r_g'. Experimental values of r_g are between 90 and 300 m for h_t between 5 and 20 m and h_r approximately 2m where h_t and h_r are, respectively, the heights of the transmitting and receiving antennae. These values are in reasonable agreement with the theoretical expression $r_g = 4h_t h_r/\lambda$ (Ground appears in the first Fresnel zone between transmitter and receiver), where λ is the wavelength of the transmitted wave.

For terminal communications κ varies from 3 to 5, where low value corresponds to rural or suburban terrains. The higher values correspond to urban environments with high-rise buildings. The plane earth model is covered by this model corresponding to a value of κ = 4 to the case where transmitter and receiver are far apart corresponding to the antenna elevations. The constant C is mainly dependent on antenna parameters such as antenna gain, elevation, etc.

4.5.4 Macro-Cell Propagation Model

The Okumara–Hata model is the most commonly used model for macro-cell coverage planning. It is used for frequency ranges 150–1000 MHz and 1500–2000 MHz. The coverage range is 1 to 20 km. The loss between the transmitting and receiving station for this model is given by,

$$L_{0-h} = A + B \log_{10} f_c - 13.82 \log_{10} h_t - a(h_r) + (44.9 - 6.55 \log_{10} h_t) \log_{10} d + L_{other} \tag{4.55}$$

where f_c = frequency in MHz, h_t = base station antenna height in metre $a(h_r)$ = a factor function of height of mobile station, d = the distance between the BS transmitter and MS transmitter in km, L_{other} = the attenuation due to land usage classes ie, the factor depends on the environment. For large city with metropolitan environment L_{other} is of the order of 3 dB and for medium city or suburban area it is 0 dB.

$$a(h_r) = (1.11 \log_{10} f_c - 0.7) h_r - (1.56 \log_{10} f_c - 0.8) \text{ [9]}$$

For a small or medium sized city (urban area):

$$a(h_r) = 8.25 (\log_{10} 1.54 h_r)^2 - 1.1, \text{for } f_c \le 200 \text{ MHz} \tag{4.56}$$

For a large city :

$$a(h_r) = 3.2(\log_{10} 11.75 h_r)^2 - 4.97, \text{ for } f_c \ge 400 \text{ MHz} \tag{4.57}$$

The value of the constants A and B varies with frequency range as given below.

A = 69.55 and B =26.16 for 150 –1000 MHz

A = 46.3 and B = 33.9 for 1000 – 2000 MHz

These variations are due to the type of terrain. This may include losses in the city area where small cells are predominant. In the forest area, there are foliage losses. Radio signal strengths depends on the type of trees, trunks, leaves, branches. It also depends on the density of the forest, height of the trees relative to the

antenna heights. Again, effects of the other natural things like water bodies, hills, mountains and glaciers etc, may affect the propagation loss. There are also the seasonal effects.

COST 231 Extension of Hata Model The Hata model was extended by the European cooperative for scientific and technical research (EURO-COST) for the frequency of 2 GHz as follows:

$$P_{L-\text{urban}}(d)\ dB = 46.3 + 33.9 \log_{10}(f_c) - 13.82 \log_{10}(h_t) - a(h_r) + (44.9 - 6.55 \log_{10}(h_t)) \log_{10}(d) + C_M \tag{4.58}$$

where a(h_r) remains same as described before, C_M is the 0 dB for medium sized city and suburbs, and 3 dB for metropolitan areas. This model is known as COST 231 extension of Hata Model and is restricted to apply in the range 1.5 GHz < f_c < 2 GHz, 30 m <h_t, 200m, 1m <h_r , 10 m and 1 km<d< 20 km> .

Example 4.14 If the received power at a reference distance d_0= 1 km is equal to 1 μWatt in the city area, find the received power at distances of 2 km from the same transmitter using Okumara–Hata model. Given height of transmitting antenna is h_t = 40 m, receiving antenna h_r = 3m, operating frequency f = 1800MHz.

Solution From Eq. (4.55) for macro cell propagation model, known as Okumara–Hata model, the loss is given by

$$L_{o-h} = A + B \log_{10} f_c - 13.82 \log_{10} h_t - a(h_r) + (44.9 - 6.55 \log_{10} h_t) \log_{10} d + L_{\text{other}}$$

As the operating frequency is 1800 MHz, A =46.3, B = 33.9, L_{other} = 3 dB

$a(h_r)$ = 3.2 (log 11.75 h_r)^2 – 4.97dB, for $f_c \geq$ 400 MHz for large city

For h_r = 3 m, a (h_r) = 2.689 dB

For d = 2 km,

$L_{\text{o-h}}$ = 46.3 +33.9 $\log_{10}$ 1800 –13.82 $\log_{10}$ 40 – 2.689 + (44.9 –6.55 $\log_{10}$ 40) $\log_{10}$ 2 + 3 = 147.87 dB

When d =1 km, $L_{\text{o-h}}$ = 133.82 dB

The received power at a distance 2 km using this model is

$$P_{r-OH} = P_0(dB) - [L_{oh}(2 \text{ km}) - L_{oh}(1 \text{ km})]$$

$$= -60 - [145.18 - 133.82] \text{ dB} = -71.36 \text{ dB} = -41.37 \text{ dBm}.$$

It is to be worth mentioning that the power received in the free space model at a distance 2 km with the same reference-transmitting signal was –36 dBm and that of two ray model was –28 dBm.

4.5.5 Micro-Cell Propagation Model:

Walsh-Ikegami [4] model is the commonly used micro-cellular propagation model in the urban environments. This model is applicable for the frequency range 800-2000 MHz and base station antenna heights up to 50 m over a distance of 5 km. The path loss for this model is dependent on two factors: Line of sight (LOS) and Non- line of sight (NLOS).

For LOS condition, $PL = 42.6 + 26 \log_{10} d + 20 \log_{10} f$ (4.58)

For NLOS condition, $PL = 32.4 + 20 \log_{10} f + 20 \log_{10} d + L_{rds} + L_{ms}$ (4.59)

In Fig. 4.34, L_{rds} is the loss due to rooftop-street diffraction and scatter loss; L_{ms} is the multi-screen diffraction loss due to the rows of buildings [16].

Distance between the mobile and base station is d in km, W_r is the width of the road, h_{bsa} and h_{bld} are the height of the base station antenna from the roof and height of the building respectively, f is the frequency in MHz.

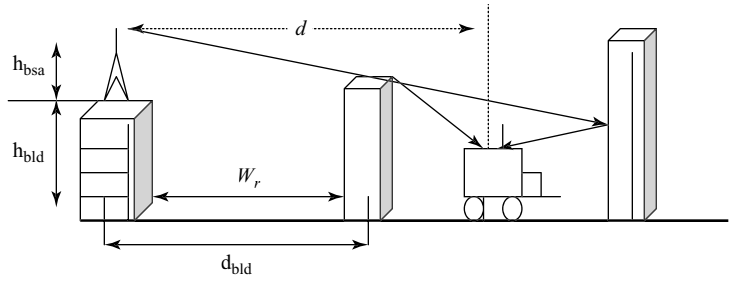


Fig. 4.34 *Microcell propagation model*

4.5.6 Shadowing Model

The free space model and the two-ray model predict the received power as a deterministic function of distance. They both represent the communication range as an ideal circle. In reality, the received power at certain distance is a random variable due to multipath propagation effects, which is also known as *fading effects*. In fact, the above two models predicts the mean received power at distance. A more general and widely used model is called the **shadowing model**.

Except for smaller objects obstructing the line of sight (LOS) path between transmitter and receiver, there may also be the presence of large mountain, hills, high rise building, etc. A mobile terminal moving behind such an object will face the effect of shadowing. Shadowing is a medium-scale effect. Field strength variations occur if the antenna is displaced over distances larger than a few tens or hundreds of metres.

Large Scale Path Loss and Shadowing Due to the diffraction, the propagation wave will enter into the geometric shadow region. At a frequency above 300 MHz, the amount of diffracted energy is low and shadows will thus be rather distinct. As the mobile terminal moves in an environment where it gets partially shadowed from direct signals and also from collections of reflected signals from different objects, the received signal at the terminal will fluctuate. This phenomenon is called **shadow fading**. Obviously, the received signal variation depends upon the position of the terminal with respect to shadowing objects. This objects may be high-rise buildings or hills, which have large physical dimensions, it may take sometime to move out the terminal from the shadowed region.

By knowing exact geometry of the terrain where the mobile is resided, shadowing gain can be calculated using various diffraction models [2]. This information is very important for cellular planning to locate the height of the contour maps and terrain to get accurate propagation prediction model. Accuracy mainly is dependent on the accurate mapping of the geographical region. Otherwise, there is always a prediction error. This prediction model is the simplest stochastic model where the distance based average gain is used. A common model is the log-normal distribution model that is characterized by the probability density function.

Log Distance Path Loss Model It is observed that both for the measurement and theoretical propagation model, the average received signal power is decreased logarithmically with distance whether in indoor or outdoor radio environment. The general expression for average large-scale path loss for an arbitrary separation of $T_x - R_x$ is expressed as follows:

$$\overline{P}_L(X) \propto (X/X_0)^{k.}$$

or in log scale, it can be written as

$$\overline{P}_L(dB) = \overline{P}_L(X_0) + 10\kappa \log_{10}(X/X_0) \tag{4.60}$$

where κ is the path loss exponent, indicates the rate at which path loss increases with distance, x_0 is the reference distance close to the transmitter that is determined from measurements and x is the T_x-R_x separation. The bar in the expressions indicate the ensemble average taken for all possible path loss values for a given value of x. Typical values of κ for different propagation environment are given in Table 4.2.

Table 4.2 *Path loss exponent values*

Environment	Path loss exponent, κ
Free space	2
Urban area cellular communication	2.7 to 3.5
Shadoewd or obstracted cellular radio	3 to 5
Indoor LOS	2.6 to 2.8
Indoor NLOS	4 to 6

The important part is to select the reference distance depending on the propagation environments. For large scale cellular coverage x_0 is taken as 1 km where as for microcell, it ranges in between 1 to 100 m, again this distance must be in far field region.

Log-Normal Shadowing Considering the presence of large obstacles in the region of transmitter and receiver with the same separation as in log-distance path loss model, the measurement of path loss at any distance x is of random value and distribute log-normally about the mean distance dependent value.

For paths longer than a few hundred meters, the received local mean power fluctuates with a lognormal distribution about the area mean power. Log normal means that the local-mean power expressed in logarithmic values.

$$\overline{P}_L(dB) = \overline{P}_L(X_0) + 10\kappa \log_{10}(X/X_0)$$

Since the location, size, and dielectric properties of the blocking objects as well as the changes in reflecting surfaces and scattering objects that cause the random attenuation are generally unknown, statistical models must be used to characterized the attenuation. The most used model for this additional attenuation is the log-normal shadowing. This model empirically derived accurately considering the variation in received power both in indoor and outdoor environments.

The probability density function (pdf) of the local-mean power is thus of the form,

$$f_{PLS}(\overline{P}_{LS}) = (1/\sqrt{2\pi}\sigma_s)\exp[(-1/2\sigma_s{}^2)(\overline{P_{LS}} - \overline{P}_{\text{Lav}})^2] \tag{4.61}$$

where σ_s is the logarithmic standard deviation of the shadowing, expressed in natural units and P_{LS} is the path loss due to shadowing effects in log scale.

The log standard deviation may be in the order of 8 to 12 dB, dependent on the terrain roughness. When more accurate and sophisticated prediction schemes are used, this value may get reduced. This log-normal fluctuation was called large-area shadowing by Marsan, Hess and Gilbert. Over semi-circular routes in Chicago, with fixed distance to the base station, it was measured to a range 6.5 dB to 10.5 dB, with a median of 9.3 dB. If the mobile moves over many kilometers, large-area shadowing reflects shadow fluctuations.

In most papers on mobile propagation, only small-area shadowing is considered: lognormal fluctuations of the local-mean power over a distance of tens or hundreds of metres are measured. Marsan *et al*. [14] reported a median of 3.7 dB for small area shadowing. Preller and Koch measured local-mean powers at 10 m intervals and studied shadowing over 500 m intervals. The maximum standard deviation experienced was about 7 dB, but 50% of all experiments showed shadowing of less than 4 dB.

Combining the distance loss and shadow fading in dB scale, we can write

$$\overline{P_{LS}}(X)(\text{dB}) = \overline{P_L}(X) + X_\sigma = \overline{P_L}(X_0) + 10\kappa \log(X/X_0) + X_\sigma \tag{4.62}$$

where X_σ is a zero mean Gaussian random variable (in dB) with standard deviation σ ($\approx$ 3.65 dB). The reference distance x_0, path loss exponent κ and the standard deviation σ, statistically describe the path loss model in a location for separation distance x between T_x - R_x. The standard deviation is a measure of the impreciseness of the terrain description. For simplicity, taking only distance dependent path loss model, the standard deviation will necessarily be large. On the other hand, for the planning of a practical network in a known environment, the accuracy of the large-scale propagation model may be refined considering the exact terrain map. This may allow more spectral efficient planning if the cellular layout is optimized for the propagation environment.

Outage Probability Under Path Loss and Shadowing In wireless systems, there is typically a target minimum received power level P_{min} below which performance becomes unacceptable. However, with shadowing the received power at any given distance from the transmitter is log-normally distributed with some probability of falling below P_{min}. Outage probability P_{out} is defined under a path loss and shadowing to be a probability that the received power at a given distance d falls below the P_{min}, and in that condition communication will be disrupted. Thus,

$$\begin{aligned} P_{out}(P_{\min,} d) &= p(P_r(d) < P_{\min}) \\ &= 1 - Q([P_{\min} - (P_t + 10\log_{10} \text{K} - 10\,\kappa \log_{10}(X/X_0))]/\sigma) \end{aligned} \tag{4.63}$$

where K = -31.54 dB, Q function is defined as the probability that a Gaussian random variable γ with mean zero and variance 1 is bigger than z,

$$Q(z) = p(\gamma > Z) = \int_Z^\infty 1/\sqrt{2\pi}\, e^{-1/2\, y^2}\, dy$$

The conversion between the Q function and complementary error function is

$$Q(z) = 1/2 \text{ erf } (z/\sqrt{2})$$

As, for example, if for a wireless communication system, the transmitted power P_t = 10 mW, $P_{\min}$ = -115 dBm, d = 200 m, d_0 =1 m, κ = 3.7, σ = 3.65 dB, the outage probability is calculated as

$$P_{out}(200\ m, -115\ \text{dBm}) = (p(P_r(200) < -115\ \text{dBm})$$

$$= 1 - Q(-115 - (P_t + 10\log_{10} K - 10\,\kappa \log_{10}(X/X_0)))/3.65)$$

$$= 1 - Q(-115 - 10 - 31.54 - 37\log_{10}(200))/3.65)$$

$$= 1 - Q((-115 + 106.7)/3.65) = 1 - Q(-2.273) = 1 - 1 + Q(2.273) = 0.012$$

An outage probability of 1% is a typical target in wireless communication. Thus from the outage probability, the concept of cell coverage area can be determined. The total area within a cell where minimum power requirements exceeds is the cell coverage area. The outage probability of the cell is defined as the percentage of the area within the cell that does not satisfy the minimum power requirement $P_{min.}$

We will now compute **cell coverage** under path loss and shadowing. The percentage area within a cell where the received power exceeds a minimum threshold is obtained by increasing the incremental area dA with radius r from the centre of the BS of the cell as shown in Fig.4.35

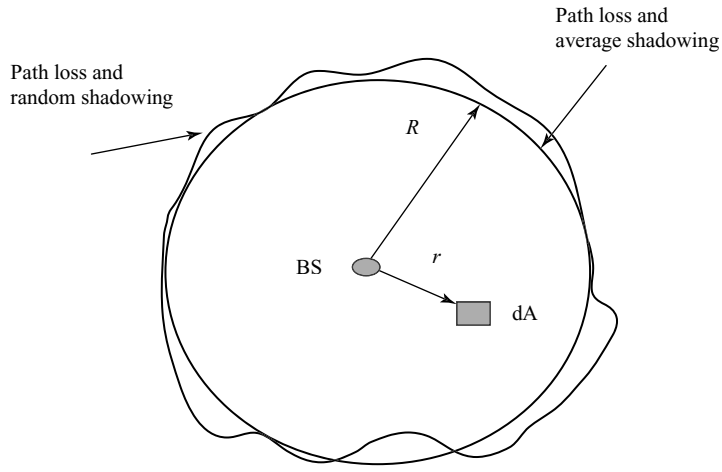


Fig. 4.35 *Received constant signal power contour of a cell*

Let us consider that the received power at dA be $P_r(r)$ considering the combined path loss and shadowing effect within the multipath. The total area within the cell that exceeds the minimum power requirement is obtained by integrating the overall incremental areas where the minimum power exceeds.

$$C = E\left[\frac{1}{\pi R^2}\int_{cell\ area}[P_r(r) > P_{\min} \text{ in } dA]dA\right]$$

$$\text{Let, } P_A = p(P_r(r) > P_{\min}) \text{ in } dA$$

$$C = \frac{1}{\pi R^2}\int_{cell\ area} P_A dA = \frac{1}{\pi R^2}\int_0^{2\pi}\int_0^R P_A r dr d\theta \tag{4.64}$$

The outage probability is thus defined as the percentage of area within the cell that does not meet its minimum power criterion. So, $\boldsymbol{P_{out}}$ **= 1-C**.

The outage probability can be defined as in Eq. (4.63) with x=r,

$$\begin{aligned} P_{out}(P_{\min,}d) &= p(P_r(d) < P_{\min}) \\ &= 1 - \mathrm{Q}([\mathrm{P_{min}} - (\mathrm{P}_t + 10\log_{10}\mathrm{K} - 10\,\kappa\log_{10}(\mathrm{r}/X_0))]/\sigma = 1 - \mathrm{P_{out}}(\mathrm{P_{min}}, \mathrm{r}) \end{aligned} \tag{4.65}$$

Locations within the cell with received power less than P_{min} is the outage location where mobile fails to receive signal.

Now combining Eqs. (4.64) and (4.65),

$$C = \frac{2}{R^2}\int_0^R rQ\left(a + b\ln\frac{r}{R}\right)dr, \tag{4.66}$$

where $a = (P_{\min} - \overline{P_r}(R))/\sigma, b = 10\,\kappa\log_{10}(e)/\sigma$

$\overline{P_r(R)} = (P_t + 10\log_{10}K - 10\,\kappa\log_{10}(R/X_0))$. power received at the cell boundary at a distance R.

Example 4.15 The power measurement at the receiver were taken at a distance of 100 m, 200 m, 1 km and 2 km from a base transmitter station. The corresponding measured values at these distances are –0 dBm, -25 dBm, -35 dBm and –38 dBm respectively. For another sets of measurement the power measurement are taken as –0 dBm, -30 dBm, -40 dBm and –50 dBm respectively. It is assumed that the path loss for these measurement follows the Log-Normal shadowing model as given by Eq. (4.62). The reference distance $x_0 = 100$ m. Then

(a) Find the path loss exponent κ for the minimum mean square error (MMSE) estimate in the two cases. Then find the average κ.
(b) Calculate the standard deviation of shadowing about the mean value for the two cases.
(c) Using this model, estimate the received power at 1 km, 2 km for two cases.
(d) In terms of Q function, predict the likelihood that the received signal level at 1 km will be greater than –35 dBm for the two values of κ.
(e) What will be the probability that the received signal level will be less than threshold –35 dBm for the two values of κ?
(f) Comments on the results obtained with respect to κ.

Solution

(a) First, we need to calculate the value of κ from the given data.
The reference distance $x_0 = 100$ m, $P_r(100 \text{ m}) = 0 \text{ dBm} = P_0(\text{dB})$

$$P_r(X) = P_0(\text{dB})(\text{Reference factor}) - \text{PL(dB)(loss factor)}$$

The governing equation for the loss calculation is,

$$\overline{P_{LS}}(x)(dB) = \overline{P_L}(x) + X_\sigma = \overline{P_L}(x_0) + 10\kappa \log_{10}(x/x_0) + X_\sigma$$

Let at a distance x_i the received power is P_{ri} and the estimation of this power is P'_{ri}.
Considering all the four measurements, the MMSE can be calculated as for the first sets of values as:

$$P_n = \sum_{n=1}^{4} [P_{ri} - P'_{ri}]^2$$

$$= [0-0]^2 + [-25-(0-10\times\kappa\log(200/100))]^2 + [-35-(0-10\times\kappa\log(1000/100))]^2$$

$$+[-38-(0-10\times\kappa\log(2000/100))]^2$$

$$= 625 - 150\kappa + 9\kappa^2 + 1225 - 700\kappa + 100\kappa^2 + 1445 - 988\kappa + 169\kappa^2$$

$$= 278\kappa^2 - 1838\kappa + 3294$$

To get the optimum κ, we differentiate P_n with respect to κ and equating to zero
We get, $\kappa_1 = 3.3$ (the required path loss exponent)
For the second sets of values as

$$P_n = [0-0]^2 + [-30-(0-10\times\kappa\log(200/100))]^2 + [-40-(0-10\times\kappa\log(1000/100))]^2$$

$$+[-50-(0-10\times\kappa\log(2000/100))]^2$$

$$= 278.3201\kappa^2 - 2281.6\kappa + 5000$$

To get the optimum κ, we differentiate P_n with respect to κ and equating to zero
We get, $\kappa_2 = 4.098$ (the required path loss exponent)
Averaging the two $\kappa = (4.098 + 3.3)/2 = 3.67$

$$P_n(\kappa = 3.3) = 278 \times 3.3^2 - 1838 \times 3.3 + 3294 = 255.6$$

$$P_n(\kappa = 4.098) = 278.32 \times 4.098^2 - 2281.6 \times 4.098 + 5000 = 323.999 = 324$$

(b) The standard deviation of shadowing $\sigma_S = \sqrt{(P_n/4)_{\kappa=3.3}} = \sqrt{(255.6/4)} = 8$ dB

For the second cases $\sigma_S = \sqrt{(P_n/4)_{|\kappa=4.098}} = \sqrt{(324/4)} = 9$ dB

(c) The received power estimation at a distance 1 and 2 km with ($\kappa = 3.3$) obtained as

$$P_r(= 1 \text{ km}) = P_0(\text{dBm}) - 10 \times \kappa \log(1000/100)\ (\text{dB})$$
$$= 0 - 33 = -33 \text{ dBm}$$

$$P_r(= 2 \text{ km}) = P_0(\text{dBm}) - 10 \times \kappa \log(2000/100)\ (\text{dB})$$
$$= 0 - 42.94 = -42.94 \text{ dBm}$$

The received power estimation at a distance 1 and 2 km with ($\kappa = 4.098$) obtained as

$$P_r(= 1 \text{ km}) = P_0(\text{dBm}) - 10 \times \kappa \log(1000/100)\ (\text{dB})$$
$$= 0 - 40.98 = -40.98 \text{ dBm}$$

$$P_r(= 2 \text{ km}) = P_0(\text{dBm}) - 10 \times \kappa \log(2000/100)\ (\text{dB})$$
$$= 0 - 53.31 = -53.31 \text{ dBm}$$

Considering X_σ the added Gaussian random variable with zero mean and standard deviation 8 dB and 9 dB for the two path losses the resultant received power will be modified by adding X_σ dBm.

(d) For likelihood ratio, determination of threshold is the most important part. The probability that the received signal level will be greater than –35 dBm is given by the Q function. This is illustrated in Fig. 4.36

Here the threshold is γ= -35 dBm, only valid when x = 2 km. The likelihood ratio can be defined as

$$P_r[P_r(X) > -35 \text{ dBm}) = Q((\gamma - \overline{P_r(X)}/\sigma)) = Q((-35 + 42.94)/8) = Q(0.9925)$$

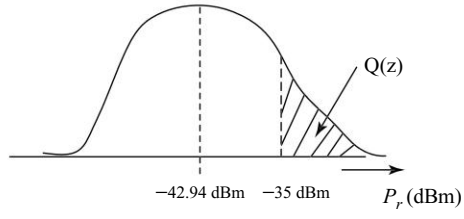


Fig. 4.36 *Received signal power for Problem 4.15*

For $\kappa = 4.098$
the power received is greater than the defined threshold at x = 1 km and x = 2 km.

$$P_r[P_r(X) > -35 \text{ dBm}) = Q((\gamma - \overline{P_r(X)}/\sigma)) = Q((-35 + 40.98)/9) = Q(0.664) = 0.25$$

$$P_r[P_r(X) > -35 \text{ dBm}) = Q((\gamma - \overline{P_r(X)}/\sigma)) = Q((-35 + 53.31)/9) = Q(2.0344) = .0227$$

(e) The probability that the received signal level will be less the threshold –35 dBm is given by

$$P_r[P_r(X) < -35 \text{ dBm}) = Q((P_r(X) - \gamma)/\sigma)) = Q((-42.94 + 35)/8)$$
$$= Q(-0.9925) = 1 - Q(0.9925) = 1 - .184 = 0.816$$

$$Q((-40.98 + 35/9) = 1 - Q(0.664) = 1 - 0.25 = 0.75$$

$$Q((-53.31 + 35/9) = 1 - Q(2.0344) = 1 - .0227 = 0.9773$$

(f) As the loss increases the propagation path loss constant κ also increases. This is also observed from the calculation. With the greater κ, the standard deviation of the loss is increased which is more likely in a lossy medium.

Example 4.16 In a measurement process of received signal within an indoor environment, the measured data follows the distance dependent mean power law that satisfy log-normal distribution as $P_r(d) \propto d^{-3.3}$. At a reference distance d_0 = 1m from the base transmitter, the received signal is 1 mW. At a distance of 10 m, it is found that 11.5% of the measurements were stronger than the –28 dBm.

(a) Find the standard deviation σ, for the path loss model at a distance d = 10m

(b) Find σ, when the path loss constant is 3.8 keeping other parameters same

Solution

(a) We have the path loss exponent $\kappa = 3.3$

Mean power at a distance d is

$$\overline{P_r(d)}(\text{dBm}) = P_r(d_0) \text{ dBm} - 10 \times \kappa \log d/d_0$$
$$= 0 - 33 = -33 \text{ dBm}$$

For the problem definition, the threshold power level γ= -28 dBm

$$P_r[P_r(d) > \gamma] = Q((\gamma - \overline{P_r(d)}))/\sigma = 11.5\%$$

$$\Rightarrow Q((-28 + 33)/\sigma) = 11.5\%$$
$$\Rightarrow Q(5/\sigma) = 0.115$$
$$\Rightarrow 5/\sigma = 1.20$$

- $\sigma = 4.16$ dB
- For $\kappa = 3.8$

$$\overline{P_r(d)} \text{ (dBm)} = P_r(d_0) \text{ dBm} - 10 \times \kappa \log d/d_0$$
$$= -38 \text{ dBm}$$

$$P_r[P_r(d) > \gamma] = Q((\gamma - \overline{P_r(d)}))/\sigma = 11.5\%$$
$$\Rightarrow Q(10/\sigma) = 0.115$$

$$\Rightarrow 10/\sigma = 1.2$$
$$\Rightarrow \sigma = 8.33 \text{ dB}$$

Example 4.17 The propagation path loss model restricts the cell coverage area in a cellular network. Consider a log-normal path loss model, the received signal power at a distance $d > d_0$ is

$$P_r(d)(\text{dB}) = P_r(d_0)\ \text{dBm} - 10\,\kappa \log_{10}(d/d_0) + X_\sigma(\text{dB})$$

Where all the parameters have their usual meaning. X_σ(dB) is a random variable uniformly distributed over [-b, b]. The cell coverage is defined as the service area of a base station over which the signal power at a mobile station is greater than a threshold γ with a probability of b. Find the expression for evaluating the cell coverage area.

Solution The distribution function of X_σ is shown below.

The cell coverage is defined as the service area of a base station in which the signal power (from the BS) received at a receiver is larger than γ with the probability of b.

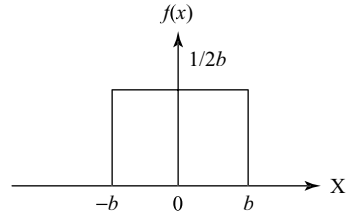


Fig. 4.37 *Random variable* X_σ

Now, γ (dBW)=10 $\log_{10}$ γ and $P_r >= 10 \log_{10} \gamma$

$$P_r(d_0) = \text{dBW} - 10\,\kappa \log_{10}(d/d_0) + X_\sigma\ (\text{dB}) \geq 10 \log_{10}\ \gamma$$

or, $\gamma(\text{dBW}) \geq 10 \log_{10}(d/d_0)^\kappa \gamma - P_r(d_0)$ dBW

or γ (dBW) $\geq \beta$ (say)

$\beta = 10 \log_{10}\ (d/d_0)^\kappa \gamma - P_r(d_0)$ dBW

$$\text{Now, } p(\gamma\ (\text{dBW}) \geq \beta) = \int_\beta^\infty f(x)\,dx \tag{i}$$

$f(x)$ is shown as in Fig. 4.37.
Let $d = r$ and $\beta = \beta(r)$ a function of r.

$$p(\gamma\ (\text{dBW}) \geq \beta) = \int_\beta^\infty f(x)\,dx = u\ (r)\ (\text{say}) \tag{ii}$$

$b = 1/A \iint_A u(r)\,dA$, where A represents a circle of radius $R > d_0$

$$= \frac{1}{\pi(R^2 - d^2{}_0)} \int_{\theta=0}^{2\pi} \int_{r=d_0}^{R} u(r) r\,dr\,d\theta$$

$$= 2/(R^2 - d^2{}_0) \int_{d_0}^{R} u(r) r\,dr \tag{iii}$$

So, given the value of b, the coverage area can be determined from Eqs. (ii) and (iii).

Example 4.18 A WSSUS channel has a multipath delay spread T_m =1 s and Doppler spread of f_d = 0.02 Hz. The total channel bandwidth at band pass available for signal transmission is W_s = 10 Hz. To reduce the effects of ISI, the signal designer selects a symbol duration T_s =10 s.

(a) Determine the coherence bandwidth and the coherence time .
(b) Does the channel exhibit frequency selective fading? Explain.
(c) Does this channel exhibit slow and fast fading? Explain.
(d) Determine the transmission data rate of the system.

Solution

(a) Coherence time, $t_c = 1/f_d = 1/.02 = 100/2 = 50$ Sec ,
Coherence BW, $f_c = 1/T_m = 1/1\ s = 1$ Hz
(b) As f_c (1 Hz) << channel BW (10 Hz), the channel fading is frequency selective
(c) Again as $T_s << t_c$, thus the channel is slow fading.
(d) Data rate $R_s = 1/\ T_s = 1/10$ s = 0.1 bps.

Example 4.19 A wireless channel is specified by the time-variant channel impulse response,

$$h(\tau,t) = (1-\tau/t)\cos(\omega t+\varphi_0), 0 \le t \le T,$$

where $T = 0.1$ms, $\omega = 10\ \pi$, and φ_0 lies between $[-\pi$ to $+\pi]$ is a constant.

(a) Determine the channel time variant transfer function
(b) Given that the channel input signal is

$$X(t) = \begin{cases} 1, & 0 \le t \le T_s, \\ 0, & \text{otherwise} \end{cases}$$

Find the channel output which depends on T.
(c) For continuous digital transmission with symbol duration T_s, find the relation between T and T_s, if the channel is considered frequency selective fading. What will be the value of T, to make the channel non-frequency selective.

Solution (a) Time variant transfer function,

$$H(f,t) = FT_\tau\, h(\tau,t) = \int_0^T h(\tau,t)e^{-j2\pi f\tau}d\tau$$
$$= \cos(\omega t+\varphi_0)[1/j2\pi f + (1-e^{-j2\pi fT})/(2\pi f)^2]$$

(b) Channel output $r(t) = \int_0^T h(\tau,t)\times(t-\tau)\,d\tau, \qquad 0 \le t \le T_s$
$$= 0, \qquad T \le 0$$
$$r(t) = \int_0^t (1-\tau/T)\cos(\omega t+\varphi_0)d\tau, \qquad 0 < t < T_S$$
$$= \cos(\omega\, t+\varphi_0)\,[t - t^2\ /\ 2\,T]$$
$$r(t) = \cos(\omega t+\varphi_0)\int_t^{t+T_S}(1-\tau/T)\,dt \qquad t \ge T_S$$
$$= T_S\cos(\omega t+\varphi_0)[1-T_S/2T - t/T]$$

(c) To answer this part, we have to find delay PSD,

We know, $\Phi_h(\tau;\Delta t)\delta(\tau_1-\tau_2) \triangleq \Phi_h(\tau;\Delta t) = 1/2E(h^*(\tau_1;\tau)h(\tau_2;t+\Delta t)$ and the scattering function of the random channel is defined as the Fourier transform of $\Phi_{\rm h}(\tau;\Delta t)$ with respect to Δt as

$$S_c(\tau,\gamma) = \int_{-\infty}^{\infty} \Phi_{\rm h}(\tau;\Delta t)e^{-j2\pi\gamma\Delta t}d\Delta t,$$

where γ is the Doppler frequency shift. The **Power Delay Profile** (PDP) is defined as the auto correlation function with Δt =0, and is denoted as $S(\tau)=\Phi_{\rm h}(\tau;0)$.

$$\varphi_h(\tau) = FT_{\Delta\tau}\left\{1/2E[h^*(\tau,t)h(\tau+\Delta\tau,t)]\right\} =$$

FT of autocorrelation function = PSD as a function of propagation delay τ

$$= FT_{\Delta\tau}\left\{1/2E[(1-\tau/T)\cos(\omega t+\varphi_0)\times(1-(\tau+\Delta\tau)/T)\ \cos(\omega t+\varphi_0)]\right\}$$

$$= FT_{\Delta\tau}\ [(1/4)\{1-(\tau+\Delta\tau)/T-\tau/T\ +\tau(\tau+\Delta\tau)/T^2\}\delta(\tau)]$$

Thus $S(\tau) = \Phi_h(\tau;0)$

$$= \int 1/4[1-(\tau+\Delta\tau)/T-\tau/T+\tau(\tau+\Delta\tau)/T^2]\ e^{-j2\pi f(\Delta\tau)}d(\Delta\tau)_{|\Delta t=0}$$

$$= (1-2\tau/T+\tau^2/T^2)/4$$

The mean delay is

$$\bar{\tau} = \frac{\int_0^{\infty}\tau S(\tau)d\tau}{\int_0^{\infty}\tau S(\tau)d\tau} = T/4$$

The root mean square delay is

$$\sigma_\tau = \sqrt{\frac{\int_0^{\infty}(\tau-\bar{\tau})^2 S(\tau)\,d\tau}{\int_0^{\infty}S(\tau)\,d\tau}} = 0.096\,T \approx 0.1\,T$$

The multipath delay spread $T_m \approx \sigma_\tau = 0.1\ T$,
The symbol interval is T_s. For Frequency selective fading $T_s < T_m$ or $T_s < 0.1\ T$

This implied $T > 10\ T_s$. As T =0.1 ms, T_s should be less than 0.01 ms for frequency selective fading. If the fading is frequency non-selective or flat, then $T_s >> T_m$, or $T_s >> 0.1T >> 0.1 \times 0.1$ ms = 0.01 ms.

4.6 MOBILE COMMUNICATION ANTENNAS

Antennas for mobile communication are of different type, depending on their usage. To create radio links between one base transmission system (BTS) and the others, parabolic antennas are the most frequently used, whereas to create links between cellular phones and BTS, dipole arrays are generally used. So far we have studied about the cellular communication systems and design fundamentals, we understand that antennas play a great role on the system regarding the capacity enhancement and other aspects like multipath propagation. Multipath propagation is the dominant effect in urban mobile communication as fading of the signal is a very common phenomenon. Fading characteristics are highly dependent on the nature of local environment, time of a day and other factors. There is a demand to cope up the ambient conditions. This leads to the requirements of new innovative antennas both for base transmitter and mobile handset.

To create antenna hardware, one must consider the propagation and radio transmission characteristics [18]. The arrangement of the antenna pair in a reception diversity system virtually determines the transmission characteristics of the radio channels. Today, antennas used in mobile communication systems have been recognized as the critical elements that can enhance or constrain the system performance. Requirements for antennas depend on the types of mobile systems such as land, maritime, aeronautical and satellite mobile systems. Antennas used in base stations have different role from those used in mobile stations. So, design specifications are varied in two cases. The electrical performance and mechanical configuration of a base station antenna mainly depends on the size and shape of the service area.

In general, mobile terminal antennas are small, lightweight and low profile. It requires an omnidirectional radiation pattern in the horizontal plane and also has to be robust enough to encounter the mobile environmental hazards. In the early day, wire antennas such as whip antennas, monopole antennas and inverted –L antennas were commonly used in the vehicle and mobile unit. With the technological development of mobile communication, antenna technology has also been much progressed. The rapid growth of the personal use of mobile handset for cellular communication, development of small mobile terminals and small size radiating system are also required.

The radiation pattern of the base station antennas in cellular land mobile systems is not always omni-directional, but designed to conform to the pattern in the horizontal plane and tilted downward pattern in the vertical plane to minimize the co-channel interference. Table 4.3 provides some of the typical antennas used in the mobile communication systems.

Table 4.3 *Antennas for mobile communication systems*

Frequency band	Mobile Station		Base station	
	Antennas	**Requirements**	**Antennas**	**Requirements**
Mobile telephones 800 MHz	λ/4 monopole λ/2 sleeve dipole Printed dipole	Omni-directional pattern in horizontal plane Low elevation angle in vertical plane, space diversity	Collier array Array of dipoles Broadside array Corner reflector with two dipoles, for macro-cellular array antennas are arrays or Yagi antennas, for micro or pico cells patch antenna arrays are very common	Omni-directional pattern in horizontal plane Tilted pattern in vertical plane Low sidelobe in vertical plane, both space and polarization diversity
Portable 800 MHz	λ/4 monopole λ/4 whip normal mode helix Planar IFA	To mount antenna in a small space Inclusion of the Body or portable unit as antenna system Space diversity	Same as given above	Same as given above

Some care in designing antenna for mobile unit is very much required. The gain of the antenna should be optimum such that it is so high that base transmitter power can be lower. To avoid interference problems between cells, the transmission power used from each BTS is relatively low, according with the limited size of the area (or cell) to be covered by each BTS. Use of smaller battery and battery charging time requirement need to be considered. Again the power should not be so high so that the electromagnetic energy may affect the sensitive body parts like eyes, brain, etc. In the urban areas, diversity reception technique may be needed for both base station and mobile station antennas to encounter the effect of multipath propagation. The problem due to delay spread in the incident waves can be mitigated by angle or directivity diversity with the suitable techniques of digital communications.

In this section, some general idea will be provided in the designing of antennas both for base stations and mobile stations.

4.6.1 Land Mobile Antenna Systems

Land mobile communication systems employ base stations (BS) and mobile stations (MS). Different design criteria are also employed for these systems. Base station antenna is sometimes known as **cell site antennas**,

a common term generally used in the cellular wireless networks. A cell site is composed of a tower or other elevated structure for mounting antennas, and one or more sets of (transmitter/receivers) transceivers, digital signal processors, control electronics, timing synchronization systems, power control system, etc.

In designing the overall antenna system both the electrical and mechanical characteristics and trade-off policies for the performance and cost are to be considered [18]. It is also important to assess installation cost apart from the antenna design as a single unit. Sometimes the instalation cost may be larger than the cost of the antenna. Figures 4.38 and 4.39 are the design requirement for base station and mobile station antennas respectively. Mobile station antennas are categorized as antennas for mobile mounting unit on a vehicle, and antennas for portable mobile handset.

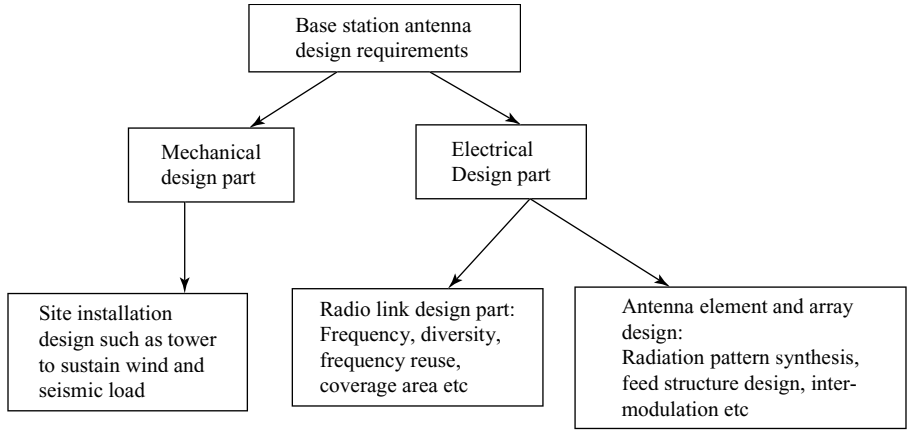
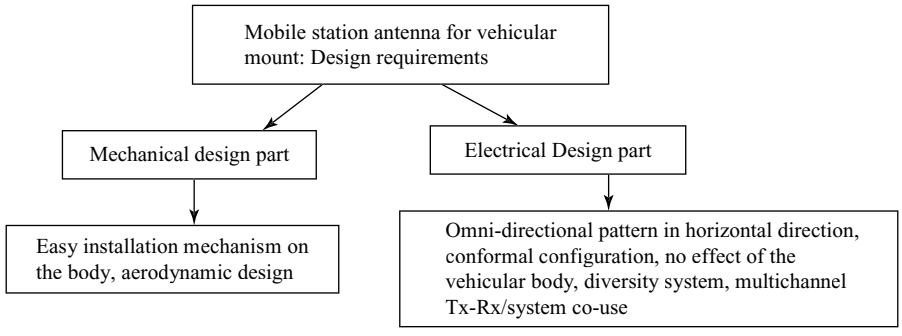
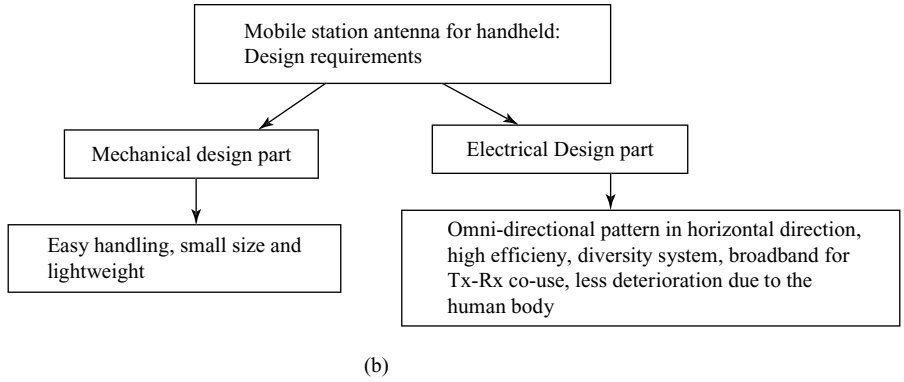


Fig 4.38 *Design requirements for base station antennas*

In cellular communication, in order to communicate with the mobile units, the base station antenna needs to radiate uniformly within the service area with sufficiently high gain. To cover the total service area, the horizontal pattern has to be omnidirectional. The possibility of achieving high gain, is to narrow the vertical beam width. To achieve this, vertical arrays of linear array are used (gain of the order of 7 to 15 dBd, with respect to dipole antenna). Again, for multiplexing the radio channel to the many users, wideband multi-channels are required. To use the frequency reuse of the cellular wireless systems, base station antenna radiation patterns plays important role for overall efficiency. The beam tilting and beam shaping is the requirement for frequency reuse. To combat the effect of fading in the propagation environment, diversity of signal reception is a must. So, both the mobile and base station antenna systems need to handle diversity schemes [19].



(a)



(b)

Fig. 4.39 *(a) Design requirement for vehicular mounted mobile station antenna, (b) Mobile handset antenna*

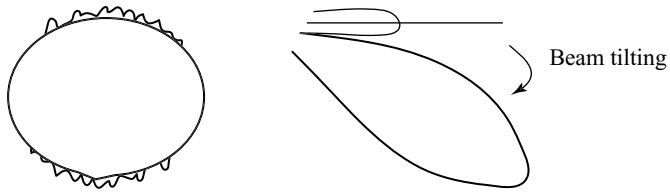


Fig. 4.40 *Typical base station antenna radiation pattern*

Types of Base Station Antenna Depending on the size and shape of service area and total number of cells, frequency reuse pattern, number of channels, the base station antenna configuration depends. For smaller service area low power antenna is required. Again, corner reflector antenna as shown in Fig. 4.41 is used if the required coverage area is restricted to a limited zone such as sectored zone within a cell. For wider area coverage, linear array antenna, which has large directivity in vertical plane, is used. A typical array antenna is shown in Fig. 4.41. The feed system should have low loss. For frequency reuse in a cellular environment the desired signal to undesired signal (D/U) ratio becomes more important than to have a high gain antenna. So, main beam tilting electronically or mechanically has been widely adopted. With proper antenna array synthesis and optimization techniques, side lobe suppression may be achieved that also helps in frequency reuse and planning. Space diversity between two antennas with a separation of 5 –10 λ in horizontal plane is generally used for diversity of reception to reduce the fading effect. The correlation coefficient between the horizontal diversity antenna should be less than 0.6, for this to achieve the antenna spacing is larger than 5λ in the urban area, and more than 20λ in suburban area. Again, the correlation coefficient is dependent on the antenna height. Generally, vertically spaced diversity antenna is used for mobile station antennas [20].

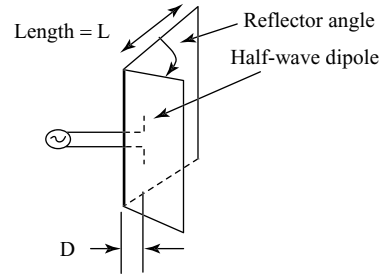


Fig. 4.41 *Geometry of a corner reflector antenna element (a base station antenna is vertically arrayed of this element)*

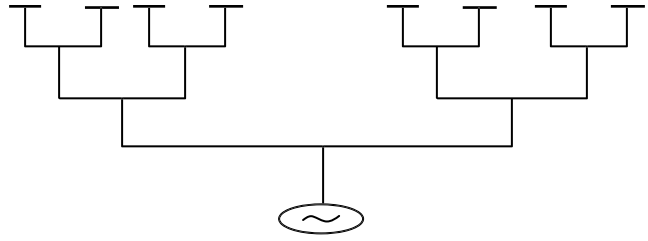


Fig. 4.42 *Typical parallel feed array antenna*

Types of Mobile Station Antennas The design objectives of mobile station antennas are lightweight, small sized, and economic at the same time with required quality of service for speech and outage probability within the cell coverage area. Within a cell, the movement of mobile station is random. For vehicular antennas omnidirectional pattern is required, particularly in a suburban areas where the base station antenna and mobile station antenna are in line of sight. Otherwise, the received signal level would vary at different location of the mobile unit. The 0 –50° angular distribution of arrival of the radio signal is common in urban and suburban area, though the mean angular distribution depend on the propagation medium. Maximum radiation in the horizontal direction is required. Vertical polarization is usually used in mobile station to develop broadband omni-directional antennas. Whip antenna or dipole antenna are commonly used.

For portable mobile telephone handset, the transmitting power is the main concern for limited battery power. The antenna gain is less than the vehicular antennas. Effective antenna gain is the important to have radiation efficiency due to the proximity of human body. The head of the user absorb and scatter the radio energy from the antenna. The azimath coverage is not uniform if the head is close to the radiator. So, variation of radiation pattern and polarization would be there as the mobile handset is randomly directed while talking. Cellular phone antennas are the relatively large bandwidth (10%), and light weight. The widely used mobile handset antennas are:

a. Sleeve dipole, b. Helical antenna c. λ/4 whip, d. Dipole helical combination, and e. Microstrip printed antenna.

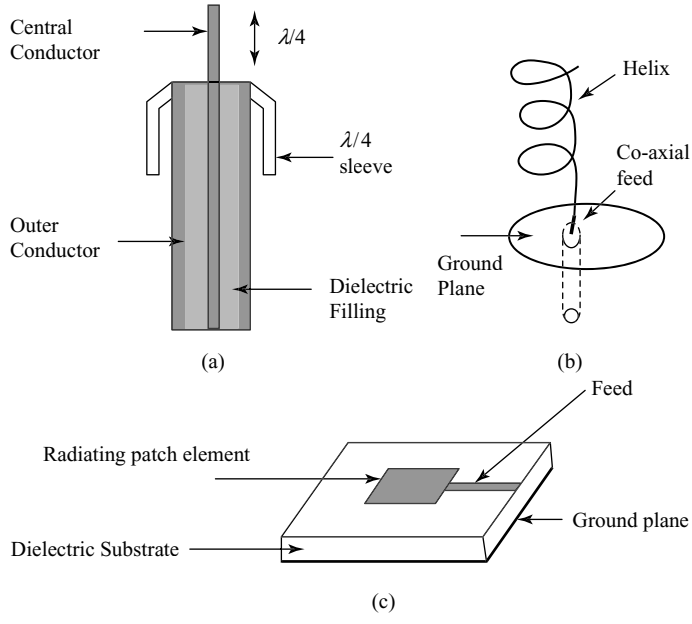


Fig. 4.43 *(a) Cross section of dipole sleeve antenna, (b) Helical antenna, (c) Microstrip patch antenna*

To obtain broad bandwidth characteristics, antenna designers transformed the horizontal element from a wire to a plate as indicated in Fig. 4.44 and the planar inverted-antenna (PIFA) was introduced [21].

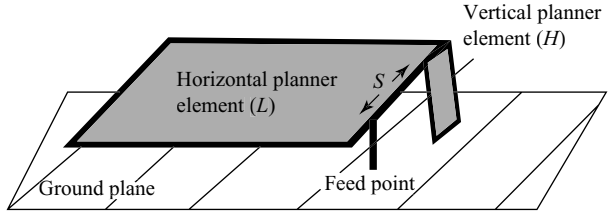


Fig. 4.44 *PIFA antenna structure used for mobile handheld antenna*

The PIFA is widely used in mobile handheld devices. It is a self-resonating antenna with purely resistive impedance at the frequency of operation. This makes it a practical candidate for mobile handheld design since it does not require a conjugate circuit between the antenna and the load, reducing both cost and losses. The evolution of the handset antenna designs from a monopole to the PIFA indicates that the essential component of a handset antenna is the "wire". The patch(s) slot(s), and stub(s) are only used to compensate for the mismatch and improve the radiation characteristics. Notice that at the megahertz frequency range, the current flowing on the surface of a conductor no longer has a uniform distribution due to the skin effect, but it is confined to a relatively small area. Therefore, the effective cross-sectional area of the conductor is smaller than the actual dimension, which is helpful for making smaller antenna in handheld devices.

Summary

Propagation media plays major challenges for wireless communications and influences greatly to characterize the channel. This chapter describes the very important aspects of propagation effects under various conditions while mobile users reside under the coverage of cellular communication. Different fading phenomena are clearly described along with the appropriate modeling. The principles of channel modeling are based on two techniques, deterministic and Empirical models and the second type is the stochastic modeling. Empirical models are channel models that depend on observation and measurement data of a particular location. Stochastic models use the first and second order statistical properties of the channels impulse response to characterize the channel behavior. Sufficient examples and illustrations are given for the proper understanding of the subject of channel characteristics and propagation models. Finally, the important part of any communication system is the proper antenna design. The mobile communication through cellular networks needs the desired radiation patterns for base station and mobile station. The objective of designing base and mobile station antennas and their types has been discussed briefly.

References

[1] Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, 2nd ed., McGraw-Hill, New York, 1984.

[2] Clarke, R.H., *A Statistical Theory of Mobile-Radio Reception*, Bell Syst. Tech. J.,Vol. 47, pp. 957–1000, 1968.

[3] Gans, M.J., *A Power Spectral Theory of Propagation in the Mobile Radio Environment*, IEEE Trans. Veh. Technol., Vol. VT-21, No. 1, pp. 27–38, Feb. 1972.

[4] Jakes, W.C., (ed.), *Microwave Mobile Communications*, Wiley, New York, 1974.

[5] Stein, S., *Fading Channel Issues in System Engineering*, IEEE Journal on Selected Areas in Communications, Vol. SAC-5, No. 2, pp. 68–89, Feb 1987.

[6] T. S. Rappaport, *wireless Communications, Principles and Practice*, Prentice Hall, PTR, 1996.

[7] Lee,W.C.Y., *Estimate of Local Average Power of a Mobile Radio Signal*, IEEE Transactions Vehicular Technology, Vol. 29, pp. 93–104, May 1980.
[8] Suzuki, H., *A Statistical Model for Urban radio Propagation*, IEEE Transactions onCommunications, Vol. COM-25, No. 7, 1977.
[9] Hata M., *Empirical Formula for Propagation Loss in Land Mobile Radio Services*, IEEE VT, VT-29 (3):317–325, August 1980.
[10] Misra A. R., *Fundamentals of Cellular Network Planning and Optimization 2G/2.5G.3G Evolution to 4G*, John Wiley and Sons, England, 2004.
[11] Sklar B., *Rayleigh Fading Channels in Mobile Digital Communications*, Part I and Part II, IEEE Communication magazine, pp. 90–100, 102–109, July1997.
[12] Saleh A.M., and R.A. Valenzula, *A Statistical Model for Indoor Multipath Propagation*, IEEE Journal on Selected Areas in Communications, SAC-692):128–137, 1987.
[13] Mark Jon. W., and Weihua Zhuang, *Wireless Communications and Networking*, PHI, New Delhi, 2005.
[14] Balanis C.A., *Antenna Theory,* 2nd edition. New York: Wiley, 1997.
[15] Marsen et.al, *An Integrated Propagation Mobility Interference Model*, Phy. Rev, Lett 93, 2004.
[16] Lee W.Y., *Estimate of Channel Capacity in Rayleigh Fading Environment*, IEEE Trans. On Vehicular Technology, Vol 39, Aug 1990, pp 187–189.
[17] Molisch A. F., *Wireless Communications*, Wiley India Edition, 2005
[18] Xia H., and H.L. Bertoni, *Diffraction of Cylindrical and Plane Waves by an Array of Absorbing Half Screen*, IEEE Trans. on Antennas and Propagation, Vol. 40, No 2, pp170–177, Feb. 1992.
[19] Yamada Y., Y. Ebine and K Tsunekawa, *Base and Mobile Station Antennas for Land Mobile Radio Systems*, IEICE Trans., vol E74, No. 6 1991, pp. 1547-1555.
[20] Yamada Y., K. Kagoshima and K. Tsunekawa, *Diversity Antennas for Base and Mobile Station in Land Mobile Communication Systems*, IEICE Trans. Vol. E74, No10, 1991, pp. 3202–3209.
[21] Ebine Y., T. Takahashi and Y. Yamada, *A Study of Vertical Space Diversity for Land Mobile Radio*, IEICE Trans., vol J73-B-II, No 6, 1990, pp. 286–292.
[22] Geyi W., Q. Rao, S. Ali, and D. Wang, *handset antenna design: practice and theory*, Progress In Electromagnetics Research, PIER 80, 123–160, 2008.
[23] Geyi, W., Q. Rao, and M. Pecen, *Multi-band Antenna Apparatus Disposed on a Three-Dimensional Substrate and Associated Methodology for a Radio Device*, US Patent file 32519.
[24] Geyi, W., S. M. Ali, and M. Pecen, *Multi-band Antenna and Associated Methodology for a Radio Communication Device*, US Patent file 32515.

Questions for Self-Test

4.1 The mean signal strength for an arbitrary transmitter-receiver (Tx-Rx) separation is useful in estimating the radio coverage
   a. True
   b. False

4.2 The large-scale effects determine a power level averaged over an area of tens or hundreds of metres
   a. False
   b. True

4.3 The term local-mean is used to denote the signal level averaged over a few tens of wavelengths
   a. True
   b. False

4.4 Describe the different mechanisms of multipath phenomena.

4.5 How is received power at the mobile station related with distance and path loss exponent?

4.6 What are the different path loss models available depending on the location of the receiver.

4.7 What are the main purposes of power control?

4.8 Describe the free space propagation path loss model. When is two-ray propagation model suitable? On what factors does the loss depend for this model?

4.9 What is called shadowing? Describe the different types of shadowing occur in mobile environment.

4.10 What is called Rayleigh fading? How this fading is modeled in the multipath environment?

4.11 When is Rician model considered for modeling fading in mobile environment?

4.12 Define the term EIRP? What is the path loss in dB for free space propagation model? Express the power received by a mobile station at a distance d from the transmitter in dBm.

4.13 A base transmitting antenna produces 40 W of power and applied to a unity gain antenna with 900 MHz carrier frequency. A receiving antenna with unity gain located at a distance 5 km from the base transmitter is used for the power reception. What is the received power in dBm and dBW?

Ans. –47.45 dBm

4.14 Repeat the Problem in 4.13 for the distance 10 km, 20 km. If the gain of the receiving antenna is 2, what would be the received power in dB at a distance 10 km?

Ans. –59.48 dBm, -50.46 dBm

4.15 A base-transmitting antenna with unity gain produces 50 Watt of power. The operating frequency is 900 MHz. A receiving antenna at the mobile station with gain 2 located at a distance 5 km from the base transmitter is used for the power reception. Calculate the received power at the mobile station using two-ray propagation model if the height of the base transmitter and mobile station are 40 m and 1 m respectively above the ground.

Ans. –65.9 dBm

4.16 What is the difference between the log-distance path loss model and log-normal shadowing?

4.17 The received signal in an pico cellular environment is found to be satisfied the log-normal distribution as $P_r(d) \propto d^{-3.5}$. At a reference distance d_0 = 1m from the base transmitter, the received signal is 1mW. At a distance of 10 m, it is found that 9.9% of the measurements were stronger than the –30 dBm.

a. Find the standard deviation σ, for the path loss model at a distance d = 10m
b. Repeat the problem for the path loss exponent 3.6, 3.7 and 3.8 and discuss the results.

Ans. –a.6.5 dB, b.7.2 dB, c. 10.4 dB

4.18 In problem 4.17, estimate

a. The received power at 10 m using the resulting model with k=3.3
b. Predict the likelihood that the received signal level at 10 m will be greater than –30 dBm. Express the answer in terms of Q function.
c. Predict the probability that the received signal level will be less the threshold –30 dBm.

Ans. –b. Q(0.769), c. 1- Q(0.769)

4.19 In a cellular communication environment, the propagation path loss model is defined by the following terms
The distance dependent average path gain (G_{av})
The shadowing gain (G_s)
The multi-path gain (G_{mp})

4.20 The received power at a point is given by the relation $P_R = P_t\, G_{av}\, G_s\, G_{mp}$
Where P_t is the transmitting power. Express the received power in dB.
The distance dependent average path loss can be modeled with some simple expression

$$G_{av}\,(\text{dB}) = 10 \log_{10} (d_0/r^{\kappa})$$

where d_0 is the reference distance =1 km, r is the location of the receiver in km, and k is the path loss exponent. Calculate the received power at a distance 2 km and 3 km for k ranging from 2 to 4. Given the

received power at a reference distance is 1 μW, the gain due to shadowing is -7 dB and due to multipath is –25 dB. Discuss the effect of k on the received power.

Hint. For d = 2 km, and $\kappa = 2$, $P_{rx} = 10^{-6} (1/2)^2 = -66$ dB = -36 dBm

4.21 In a cellular system the base station antenna height is 38 m and transmitting at 1800 MHz by an antenna with gain 10 dB. The received power is 1μW by a receiving antenna of gain 0 dB placed at a distance d and height 2 m. Then what will be the suitable distance between the transmitter and receiver if the base transmitting power is 10 W? What will be the path difference between the direct ray and the reflected ray? Find the time delay between the two rays. What will be the phase difference Δθ?

Ans. Path Difference = 0.362, time delay = 1.209 ns and Phase difference =13.64 radian

4.22 Discuss the design requirements for base station antennas. What are the different types of base station antennas are normally used?

4.23 What special things must need to consider while designing mobile handset antennas? Are there any basic design specification differences for mobile vehicular antenna with the mobile handset antennas?

4.24 How does mutipath fading is mitigated with the design of base station antennas?

4.25 Is there any relationship between the base station antenna designs with the frequency reuse of cellular system?

4.26 What are the different types of antennas used in mobile handset unit?

4.27 Why PIFA structure is gaining importance for mobile handheld devices?

4.28 Distinguish the small scale fading effect while considering NLOS components and considering LOS component (s) along with NLOS.

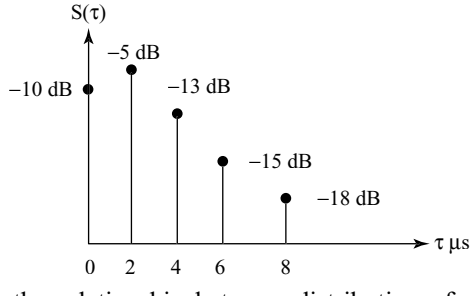
4.29 Describe the significance of coherence bandwidth and coherence time in a multipath propagation.

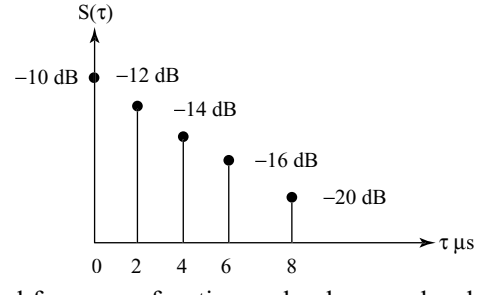
4.30 Define the following fading phenomena,
Frequency selective fading and time dispersion, Frequency dispersion and time selective fading, fast fading and slow fading, narrowband fading and wideband fading

4.31 Describe the time varying channel impulse response in wireless media. Show the magnitude plot of time varying channel impulse response.

4.32 Find the relationship of power delay profile with the maximum delay spread and rms delay spread in multipath propagation.

4.33 The discrete power profile delay for multipath transmission are shown in two figures below, show the multipath power gain, mean delay and rms delay spread for both of the cases, and comments on the results.





4.34 Establish the relationship between distribution of spaced-frequency function and coherence bandwidth.

4.35 Establish the relationship between Doppler power spectrum, Doppler spread and coherence time.

4.36 Illustrate the relationship between the four channel functions.

4.37 Define fading margin. The effect of Doppler shift of the transmitted signal is the spectral broadening in frequency domain- justify the statement.

4.38 Consider a wideband channel with multipath delay profile as given below:

$$S(\tau) = \begin{cases} e^{-\tau/.00002} & 0 \le \tau \le 20\ \mu\text{sec}. \\ 0 & \text{else} \end{cases}$$

Find the mean delay, rms delay spread and the maximum symbol rate that a linear modulated signal can be transmitted without ISI.

4.39 Calculate the outage probability for the Rayleigh fading channel with average power 2 μW and the threshold power = 3 μW. What is σ for this channel? Repeat the plot for Rician Channel with direct component = 3σ.

4.40 Plot the probability density function for the Rician distribution where the Rician factor K = −10 dB, 10 dB, and 20 dB. Comments on the results. Consider σ=.001.

4.41 Using MATLAB , plot the received signal by a mobile moving with velocity 60 km/hr. The carrier frequency is 1 GHz, phase angles are uniformly distributed, the amplitude coefficients a_n has a Gaussian distribution with zero mean and variance =.001. Consider the number of scatterer N=10 and 100. Plot the Doppler power spectrum also.

Digital Modulations for Wireless Communications

5

Introduction

The data transmission over wireless channel is digital. Today's mobile communication systems extensively use digital modulation techniques. The advancement of Very Large Scale Integration and Digital Signal Processing technology has made digital modulation more and more cost effective. *The mapping of data bits into signal waveforms for transmission is the domain of digital modulation*. At the receiver, the demodulator recovers the data bits from the received signal waveforms. The simplest modulation is the binary one where +1 bit value is mapped into one specific waveform while the –1 bit value maps into other specific waveform. The most important technique in digital communication is the M-ary modulation where group of k bits can be expressed into a symbol which again is mapped into one out of a set of $M = 2^k$ waveforms. Different modulation formats differ in the nature of waveforms that are transmitted. If T_s is the symbol duration then one symbol can be transmitted in T_s time, whereas, k number of bits (data) can be transmitted, thus data rate (bits) is higher than the signaling rate (symbol). The objective of any modulation format in digital communication is to transmit as much information as possible with certain energy level in the transmitted signal. While the channel is wireless, then Bit Error Rate (BER) is the concerning parameter to quantify the transmission quality. A desirable modulation scheme provides low BER at low received signal-to-noise ratios and performs well under wireless fading environments, requires less bandwidth and easier to implement. The existing modulation schemes may not satisfy all the schemes together, but some of them. Depending on the application requirements each modulation technique has its pros and cons.

Modulation techniques involve switching (keying) the amplitude, frequency or phase of a sinusoidal carrier in a specific fashion in accordance with the incoming data. Some of the most important digital modulation techniques are Binary Phase Shift Keying (BPSK), M-ary Phase Shift Keying (MPSK), Minimum Phase Shift Keying (MSK), Gaussian filtered MSK (GMSK), and Orthogonal Frequency Division Multiplexing (OFDM). Among these techniques, GMSK is used for GSM cellular networks and OFDM is used in WLAN and WiMAX system. In this chapter, all the standard digital modulation schemes are discussed at length. The main issue of concern in any communication system is to design optimum filter at the receiver with the consideration of modulation techniques and transmission schemes such that the error probability gets minimized in the presence of noise. There are two approaches in the receiver designing; one is the coherent receiver where phase synchronization is made at the receiver with the local oscillator used in the receiver for demodulation and the oscillator also provides carrier in the transmitter for modulation and the other is the non-coherent receiver where no phase locking is used.

5.1 ANALOGY BETWEEN VECTOR AND SIGNAL

There is a complete analogy between the signal and the vector. In the N-dimensional Euclidian space we may define a vector with lengths and angles. Figure 5.1 shows the representation of a vector in three-dimensional space.

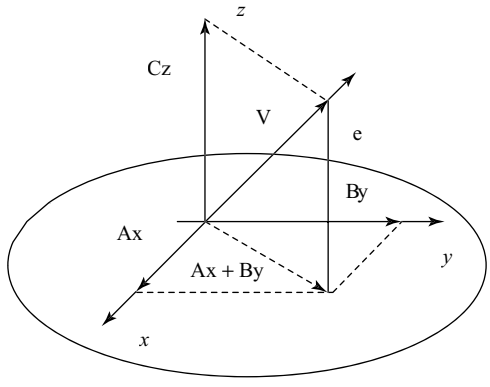


Fig. 5.1 *Three-dimensional representation of a vector*

The vector **V** has the components **Ax, By** and **Cz** along the *x*, *y* and *z* directions respectively, **x, y and z** being the unit vectors along the *x*, *y* and *z* directions respectively. Thus **V = Ax+ By + Cz**, when all the three vector components **x, y** and **z** are mutually orthogonal to each other. It is to be noted that the error **e** in the approximation is zero when **V** is approximated in terms of three mutually orthogonal vectors **x, y and z**.

The three vectors **x, y and z** represent the complete set of orthogonal vectors in the three-dimensional space. Completeness here means that it is impossible to find in this space another vector say **x1**, that is orthogonal to all three vectors **x, y and z**. Such vectors are known as *basis vectors*. If a set of vectors $\{\mathbf{x}_i\}$ is not complete, the error vector would not be zero in the approximation process. Thus in the three-dimensional case, it is generally not possible to represent a vector **V** in terms of only two vectors without an error (Example, **= Ax+By +e**).

The choice of basis vectors is not unique, but set of basis vectors corresponds to a particular choice of coordinate system. Summarizing, we can say, if a set of vectors $\{\mathbf{x}_i\}$ is mutually orthogonal, if

$$x_m . x_n = \begin{cases} 0 & \text{for } m \neq n \\ |x_m|^2 & \text{for } m = n \end{cases} \tag{5.1}$$

And if the set is complete, vector $V = c_1\mathbf{x} + c_2\,\mathbf{y} + c_3\,\mathbf{z}$, where c_i (i = 1,2,3) is the constant that can be derived as follows.

Let the vector $V = c_1\mathbf{x} + \mathbf{e}$, the component of $c_1\mathbf{x}$ and the error vector. When error is very small, then the vector V is approximated as, $V \approx c_1\mathbf{x}$. The inner or dot product of two vector is, $V.x = |V||x| \cos \theta$, where θ is the angle between the two vectors.

So, $|\mathbf{x}|^2 = \mathbf{x}.\mathbf{x}$, and the length of the component of vector **V** along **x** is $|\mathbf{V}| \cos \theta$, and then $c_1\mathbf{x} = |V| \cos \theta$ Thus, $c_1\mathbf{x}.\mathbf{x} = |V| .x \cos \theta => c_1|\mathbf{x}|^2 = \mathbf{V}.\mathbf{x}$

$$c_1 = \mathbf{V}.\mathbf{x} / \mathbf{x}.\mathbf{x}$$

In general,

$$c_i = \mathbf{V}.\mathbf{x}_i / \mathbf{x}_i . \mathbf{x}_i = \mathbf{V}. \mathbf{x}_i / |\mathbf{x}_i|^2, \qquad i = 1, 2, 3, ..., N \tag{5.2}$$

5.2 ORTHOGONAL SIGNAL SPACE

With the analogy of the vectors and signal, this concept can be extended into the orthogonal signal space. The orthogonality of a signal set $\{\mathbf{s}_i = (s_1, s_2, s_3, \ldots s_N)\}$ over the time interval $[t_1, t_2]$ is defined as

$$\int_{t_1}^{t_2} s_m(t).s_n^*(t)\,dt = \begin{cases} 0 & \text{for } m \neq n \\ E_n & \text{for } m = n \end{cases} \tag{5.3}$$

where E_n is the energy of the signal. If $E_n = 1$ for all n, the set is normalized and is called *orthonormal set*. The orthogonal set can always be normalized by dividing $s_n(t)$ by $\sqrt{E_N}$ for all n. So, the signals can be approximated over the interval $[t_1, t_2]$ by a set of mutually orthogonal signal $s_1(t)$, $s_2(t)$, ...$s_N(t)$ as,

$$s(t) = \sum_{n=1}^{N} c_n s_n(t) \quad t_1 \leq t \leq t_2 \tag{5.4}$$

when the set is complete, the error energy E_e goes to zero and the perfect equality holds. The correlation coefficients c_n (= cos θ) is expressed as

$$c_n = \cos\theta = \frac{\int_{t_1}^{t_2} s(t)s_n^*(t)\,dt}{\int_{t_1}^{t_2} s_n^{\,2}(t)\,dt} = 1/E_n \int_{t_1}^{t_2} s(t)\, s_n^*(t)\,dt \tag{5.5}$$

where $n = 1,2, \ldots N$ and * represents the complex conjugate. Thus the value of c_n lies within [–1, 1]. If the two signals are orthogonal, the coefficient is 0, for maximum similarity, it is 1 and for maximum dissimilarity it is –1.

When the set $\{s_n(t)\}$ is such that the error energy $E_e \to 0$ as $N \to \infty$ for every member of some particular class, then we say that set $\{s_n(t)\}$ is complete over $[t_1, t_2]$ for that class $s(t)$ and the set $\{s_n(t)\}$ is called the *basis functions* or *basis signals*.

Generalizing this discussion, we can write the expression for c_n for any two signals to find similarity and dissimilarity as

$$c_n = \frac{1}{\sqrt{E_1\, E_2}} \int_{t_1}^{t_2} s_1(t)\, s_2^*(t)\,dt \tag{5.6}$$

5.3 GEOMETRIC REPRESENTATION OF TRANSMITTED SIGNALS

In any communication system Additive White Gaussian Noise (AWGN) is present along with the transmitted signal. The study of the effect of the AWGN on the signal is required for taking correct decision at the receiver. Geometric representation of the transmitted signal is a very useful method to take correct decision.

In geometric representation of signals, we represent any set of M energy signals $\{s_i(t)\}$ as linear combinations of N orthonormal basis functions, where $N \leq M$. This means that given a set of real valued signals $s_1(t)$, $s_2(t)$, $\ldots s_M(t)$, each of duration T seconds , we may express $s_i(t)$ as given below:

$$s_i(t) = \sum_{j=1}^{N} s_{ij}\ \varphi_j(t), \tag{5.7}$$

For $0 \leq t \leq T$ and $i = 1,2,3, \ldots M$

$$s_{ij} = \int_0^T s_i(t)\ \varphi_j(t)\,dt, \text{ for } i = 1, 2, \ldots M,\ j = 1, 2, \ldots N \tag{5.8}$$

s_{ij} is known as the coefficient of expansion. Now the real valued functions $\varphi_1(t)$, $\varphi_2(t)$, $\varphi_3(t)$, …, $\varphi_N(t)$ are orthonormal. The word orthonormal means that

$$\int_0^T \varphi_i(t)\,\varphi_j(t)\,dt = \delta_{ij,} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} \tag{5.9}$$

where δ_{ij} is the Kronecker delta function.
The implication of Eq. (5.9) is the following:

When $i = j$, the integral is 1, it implies that each basis function is normalized to have unit energy.

When $i \neq j$, the integral value is 0, it implies that the basis functions are orthogonal w.r.t each other over the interval $0 \leq t \leq T$.

The coefficient s_{ij} is the j^{th} element of the N-dimensional vector space. The transmission system with input signal having M states ($m_i = 1, 2, \ldots M$) and the modulated energy signal $s_i(t)$ can be modeled as shown in Fig. 5.2.

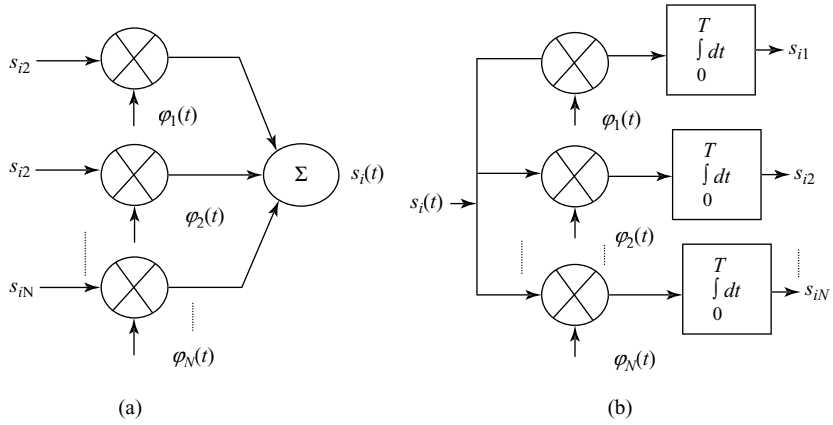


Fig. 5.2 (a) *Generation of transmitted signal* $s_i(t)$, (*b*) *Generation of the coefficient* s_{ij}

Each signal in the set $\{s_i(t)\}$ is completely defined by

$$s_i = \begin{bmatrix} s_{i\,1} \\ s_{i\,2} \\ \vdots \\ s_{i\,N} \end{bmatrix}, i = 1, 2, \ldots M \tag{5.10}$$

where $\boldsymbol{s_i}$ is called the *signal vector*. The set of signal vectors may be defined in N-dimensional Euclidian spaces for M sets of points on $\{s_i\}$ with mutually perpendicular axes $\phi_{1,}\ \phi_{2,}\ \ldots,\ \phi_{N.}$

From Fig. 5.2(b), it is seen that the generation of set $\{s_i\}$ i.e., $s_{i1,}\ s_{i2}, \ldots, s_{iN}$ is through a bank of N-product integrators or correlators with a common input and relevant basis functions. The concept of correlator is very important in the receiver design, as at the receiver the first stage is the detection of the signal. The important assumption is that the basis function used in the product integrator of the receiver are in phase i.e., coherent with the basis functions used at the transmitting end. Hence, the reception is termed as *coherent reception*.

5.4 SCHWARZ INEQUALITY

Considering two energy signals $s_1(t)$, and $s_2(t)$, the Schwarz inequality states that

$$\left(\int_{-\infty}^{\infty} s_1(t)\,s_2(t)dt\right)^2 \leq \left(\int_{-\infty}^{\infty} s_1^2(t)\,dt\right)\left(\int_{-\infty}^{\infty} s_2^2(t)\right) \tag{5.11}$$

The equality holds if and only if $s_1(t) = c\, s_2(t)$, where c is a constant.

Proof: Let us consider that $s_1(t)$ and $s_2(t)$ are expressed in terms of two orthonormal basis functions $\varphi_1(t)$, $\varphi_2(t)$ such as (consider Fig. 5.3):

$$s_1(t) = s_{11}\, \varphi_1(t) + s_{12}\, \varphi_2(t) \tag{5.12a}$$

$$s_2(t) = s_{21}\, \varphi_1(t) + s_{22}\, \varphi_2(t) \tag{5.12b}$$

where $\varphi_1(t)$, $\varphi_2(t)$ have satisfied the orthonormality conditions over the entire time interval $(-\infty, \infty)$.

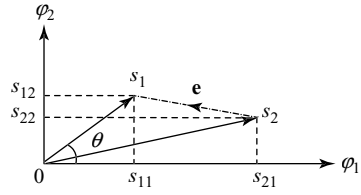


Fig. 5.3. *Vector representation of two signals $s_1(t)$ and $s_2(t)$*

When the two signals hold perfect orthogonality then $\cos\theta = 90^0$, where θ is the angle subtended by the two signals and $s_1 = c\, s_2$, c is the correlation coefficient.

$$\cos\theta = \frac{\left(\int_{-\infty}^{\infty} s_1(t) s_2(t)\, dt\right)}{\left(\int_{-\infty}^{\infty} s_1^2(t)\, dt\right)^{1/2} \left(\int_{-\infty}^{\infty} s_2^2(t)\, dt\right)^{1/2}} \tag{5.13}$$

As $|\cos\theta| \le 1$, Schwarz equality follows from Eq.(5.13)

$$\left(\int_{-\infty}^{\infty} s_1(t)\, s_2(t) dt\right)^2 \le \left(\int_{-\infty}^{\infty} s_1^2(t)\, dt\right) \left(\int_{-\infty}^{\infty} s_2^2(t)\right) \tag{5.14}$$

5.5 GRAM–SCHMIDT ORTHOGONALIZATION PROCEDURE

At this point, for Gram–Schmidt Orthogonalization procedure, we require a complete orthonormal set of basis functions. We need to prove that $\varphi_1(t)$, $\varphi_2(t)$, …, $\varphi_N(t)$ are orthonormal.

Let us consider that there are M energy signals represented as $s_1(t)$, $s_2(t)$…, $s_M(t)$. Let, the first basis function $\varphi_1(t)$ is defined as

$\varphi_1(t) = s_1(t) / \sqrt{E_1}$ where E_1 is the energy of the signal $s_1(t)$.

$s_1(t) = \varphi_1(t)\ \sqrt{E_1} = s_{11}\, \varphi_1(t)$, where coefficient $s_{11} = \sqrt{E_1}$ and $\varphi_1(t)$ has unit energy as required.

We define the coefficient $s_{21} = \int_0^T s_2(t)\, \varphi_1(t) dt$ (5.15)

A new intermediate function can be introduced as

$$g_2(t) = s_2(t) - s_{21}\varphi_1(t) \tag{5.16}$$

which is orthogonal to $\varphi_1(t)$ over the interval [0,T], and the basis function $\varphi_1(t)$ has unit energy.

We now define the second basis function $\varphi_2(t) = g_2(t) \Big/ \sqrt{\int_0^T g^2{}_2(t)\, dt}$ (5.17)

Substituting the value of $g_2(t)$, and simplifying,

$$\varphi_2(t) = \frac{s_2(t) - s_{21}\varphi_1(t)}{\sqrt{\int_0^T [(s_2(t))^2 + (s_{21}\varphi_1(t))^2 - 2.s_2(t)s_{21}\varphi_1(t)]\,dt}} = \frac{s_2(t) - s_{21}\varphi_1(t)}{\sqrt{E_2 + {s_{21}}^2 - 2{s_{21}}^2}}$$

$$= \frac{s_2(t) - s_{21}\varphi_1(t)}{\sqrt{E_2 - {s_{21}}^2}} \tag{5.18}$$

E_2 is the energy of the signal $s_2(t) = \int_0^T (s_2(t))^2\,dt$, $\int_0^T \varphi_1(t))^2\,dt = 1$, $\int_0^T s_2(t)\varphi_1(t)dt = s_{21}$

As

$$\int_0^T g_2(t)\,\varphi_1(t)\,dt = \int_0^T s_2(t)\,\varphi_1(t)\,dt - s_{21}\int_0^T [\varphi_1(t)]^2\,dt = s_{21} - s_{21} = 0$$

Hence, $g_2(t)$ is orthogonal to $\varphi_1(t)$ over the interval $[0,T]$.

$$\int_0^T \varphi_2(t)\,\varphi_2(t)\,dt = \int_0^T \frac{[\,s_2^2(t) - 2s_{21}\varphi_1(t)s_{21} + s_{21}\,(\varphi_1(t))^2\,]\,dt}{E_2 - {s_{21}}^2} = \frac{E_2 - {s_{21}}^2}{E_2 - {s_{21}}^2} = 1$$

$$\int_0^T \varphi_2(t)\,\varphi_1(t)\,dt = \int_0^T \frac{[\,s_2(t) - s_{21}\varphi_1(t)]\,\varphi_1(t)\,dt}{\sqrt{E_2 - {s_{21}}^2}} = \frac{s_{21} - s_{21}}{\sqrt{E_2 - {s_{21}}^2}} = 0$$

Hence $\varphi_2(t)$ and $\varphi_1(t)$ form an orthonormal set.

Thus,

$$s_1(t) = s_{11}\,\varphi_1(t) + 0.\,\varphi_2(t) \tag{5.19a}$$

$$s_2(t) = s_{21}\,\varphi_1(t) + \sqrt{E_2 - {s_{21}}^2}\,\varphi_2(t) \quad \text{(using Eq. 5.18)} \tag{5.19b}$$

Depending upon the above discussion, we may write the general form as given below.

$$g_i(t) = s_i(t) - \sum_{j=1}^{i-1} s_{ij}(t)\varphi_j(t) \tag{5.20}$$

where the coefficients s_{ij} are defined as

$$s_{ij} = \int_0^T s_i(t)\,\varphi_j(t)\,dt,\ j = 1,2,3,\ldots\ldots i-1 \tag{5.21}$$

It may be noted that Eq. (5.19) is a special case of Eq. (5.20) with $i = 2$. Now given the $g_i(t)$, we define the set of basis functions as

$$\varphi_i(t) = \frac{g_i(t)}{\sqrt{\int_0^T {g^2}_i(t)\,dt}} \tag{5.21}$$

Which form an orthonormal set.

So, it is proved that $\varphi_1(t)$, $\varphi_2(t)$, …, $\varphi_N(t)$ are orthonormal set. The dimensions N and M depend on one of the two possibilities.

If $s_1(t)$, $s_2(t)$, …$s_M(t)$ are linearly independent, $N = M$ and

If $s_1(t)$, $s_2(t)$, …$s_M(t)$ are not linearly independent, $N < M$ and $g_i(t) = 0$ for $i > N$.

Unlike Fourier series expansions of periodic signal, in Gram–Schmidt orthogonalization, we have not restricted the sinusoidal functions or sinc functions of time for the set of basis functions. The expansion of $s_i(t)$ is not the approximations of the terms but the exact expression where the N terms and only N terms are significant.

5.6 RESPONSE OF THE NOISY SIGNAL AT THE RECEIVER

Any signal $s_i(t)$ is represented in terms of orthonormal basis functions $\varphi_1(t)$, $\varphi_2(t)$, …, $\varphi_N(t)$ and its associated signal coefficients s_{i1}, s_{i2}, …, s_{iN}. At the receiver end if we derive the coefficients s_{i1}, s_{i2}, …, s_{iN} from the received signal, the transmitted signal would be obtained back. To get these coefficients, the received signal is multiplied by the basis functions $\varphi_1(t)$, $\varphi_2(t)$, …, $\varphi_N(t)$ and then integrated, thus obtaining the contents of $\varphi_1(t)$, $\varphi_2(t)$, …, $\varphi_N(t)$ in the received signal. This process is called *correlation reception*. During the transmission process, the signal gets corrupted with the Additive White Gaussian Noise (AWGN), so the received signal is basically the sum of the transmitted signal and the noise, which is a random process.

$$R(t) = s_i(t) + n_i(t) \text{ for } 0 \le t \le T \text{ and } i = 1, 2, \ldots, M \tag{5.22}$$

So, the output of the j^{th}, correlator is

$$R_j = \int_0^T R(t)\, \varphi_j(t)\, dt = s_{ij} + N_j, \quad j = 1, 2, \ldots N \text{ and} \tag{5.23}$$

$$s_{ij} = \int_0^T s_i(t)\, \varphi_j(t)\, dt \tag{5.24}$$

The second component N_j is the sample value of a random variable N_j that arises because of the presence of the channel noise $n(t)$ and is defined as

$$N_j = \int_0^T n(t)\, \varphi_j(t)\, dt \tag{5.25}$$

s_{ij} is the deterministic term which would have come in presence of no noise and N_j is the random term since $n(t)$ is the AWGN.

When we take the sample of $R(t)$, the result is a random process and is represented as $R'(t)$ and is expressed using Eq. (5.23) as

$$\begin{aligned} R'(t) &= R(t) - \sum_{j=1}^{N} R_j\, \varphi_j(t) = s_i(t) + n(t) - \sum_{j=1}^{N} (s_{ij} + N_j)\, \varphi_j(t) \\ &= n(t) - \sum_{j=1}^{N} N_j\, \varphi_j(t) = n'(t) \end{aligned} \tag{5.26}$$

So, the sample function $R'(t)$ depends only on the noise $n(t)$. The received signal $R(t)$ may thus be represented as

$$R(t) = R'(t) + \sum_{j=1}^{N} R_j \varphi_j(t) = n'(t) + \sum_{j=1}^{N} R_j \varphi_j(t) \tag{5.27}$$

So, the received signal is always a random (stochastic) process that is to be estimated, though the transmitted signal $s_i(t)$ is deterministic.

The transmitted signal vector, the observed signal and the noise vector can be represented as in Euclidian space as shown in Fig. 5.4. The received signal point is randomly placed about the message point, which lies on the Gaussian distributed noise cloud.

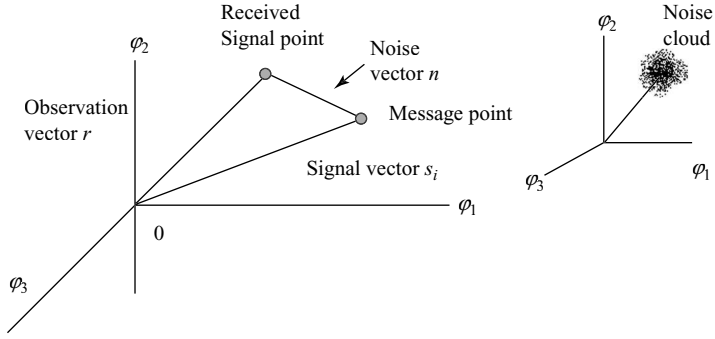


Fig. 5.4 *Received signal in the presence of noise*

5.7 MAXIMUM LIKELIHOOD DECISION RULE AND DECISION BOUNDARY

The conditional probability density functions $f_r(r|m_i)$, i = 1,2,3, …M, are basically the characterization of an AWGN channel. Its derivation leads to a functional dependence on the observation vector r, given the message symbol m_i is transmitted. At the receiver, given the observation vector r, and the requirement is the estimation of the message symbol m_i that is responsible for generating r. To emphasize this latter viewpoint, we will now introduce the idea of a likelihood function being defined as

$$L(m_i) = f_r(r \mid m_i), \quad \text{i} = 1,2,3, \ldots, M \tag{5.29}$$

In general, the likelihood ratio is expressed in logarithmic form as $\boldsymbol{l}(m_i) = \log L(m_i)$, as the probability density function is always non-negative and the logarithmic function is always monotonically increasing function.

Based on the observation vector r, the received signal $r(t)$ can be represented by a point in the same Euclidean space used to represent the transmitted signal . The received signal point is wandered around the message point in a completely random fashion as it may reside anywhere inside the Gaussian distributed noise cloud, centered on the message point as shown in Fig. 5.4.

Now we are ready to a state of signal detection problem: Given the observation vector r, a mapping is to be performed from r to estimate $\hat{m}$ from the transmitted symbol m_i in a way that would minimize the error probability during the decision making process.

Given the observation vector r, we make the decision $\hat{m} = m_i$, the probability of error in the decision process is defined by $P_e\ (m_i|r)$.

There are two types of errors associated with the reception of transmitted message m_i. If r being the received message when m_i is sent, then probability of error associated can be defined as

$$\begin{aligned} P_e\ (m_i \mid r) &= P(m_i \text{ not sent}|r) \\ &= 1 - P(m_i \text{ sent}|r) \end{aligned} \tag{5.30}$$

The decision-making criterion is to minimize the probability of error in mapping each given observation vector r into a decision. Based on this equation, the optimum decision rule is made.

$$\text{Set } \hat{m} = m_i \text{ , if } P(m_i \text{ sent}|r) \geq P(m_k \text{ sent}|r) \text{ for all } k \neq i \tag{5.31}$$

where $k = 1, 2, \ldots..M$. This is called Maximum a Posteriori Probability (MAP) rule.

In Bayesian statistics, the posterior probability of a random event or an uncertain proposition is the conditional probability that is assigned after the relevant evidence is taken into account. It is also the probability of event 'A' occurring given that event 'B' has occurred. Priori probability is the probability calculated by logically examining the existing information. It is also the probability estimate prior to receiving new information.

The condition of Eq. (5.31) may also be expressed more explicitly in terms of the *a priori* probabilities of the transmitted signals and in terms of likelihood functions. Using Bay's criterion in (5.31), and ignoring possible ties in the decision making process, the MAP rule can be restated as follows:

Set $\hat{m} = m_i$, if

$$\frac{p_k f_r(r \mid m_k)}{f_r(r)} \text{ is maximum for } k = i \tag{5.32}$$

where p_k is the *a priori* probability of transmitting symbol m_k, $f_r(r|m_k)$ is the conditional probability density function of the random observation vector R given the transmitted message symbol m_k, and $f_r(r)$ is the probability density function of random variable R which is independent of the transmitted symbol. When all the source symbols are transmitted are equally likely, then $p_k = p_i$. The conditional probability $f_r(r|m_k)$ has the one-to-one mapping to the log-likelihood function $\boldsymbol{l}(m_k)$. Accordingly the decision rule called the maximum likelihood rule can be stated as

Set $\hat{m} = m_i$ if

$$\boldsymbol{l}(m_k) \text{ is maximum for } k = i \tag{5.33}$$

It is useful to have a graphical interpretation of the maximum likelihood decision rule. If Z denotes the N-dimensional observation space for all observation vectors r, the total observation space is partitioned into M-decision regions if $m = m_i$, $i = 1,2, \ldots M$. The decision regions are $Z_1, Z_2, Z_3, \ldots Z_M$. The decision rule is restated as

Observation vector r lies in region Z_i if

$$\boldsymbol{l}(m_k) \text{ is maximum for } k = i$$

It is to be mentioned that if observation vector r falls on the boundary of two decision regions say Z_k and Z_i, then the choice between two possible decisions $\hat{m} = m_i$ and $\hat{m} = m_k$ is resolved by tossing a fair coin. The choice of the outcome will not affect the error probability as the condition of Eq. (5.31) satisfied the equality sign. The Euclidian distance between the observation point r and message point s_k is represented by $\|r - s_k\|$. Accordingly the maximum likelihood decision rule is stated as,

Observation vector r lies in region Z_i if

$$\text{the Euclidean distance } \|r - s_k\| \text{ is minimum for } k = i \tag{5.34}$$

which implies that the message point is the closest to the received signal point.
Now let us squared up the distance as

$$\sum_{j=1}^{N}(r_j - s_{kj})^2 = \sum_{j=1}^{N} r_j^2 - 2\sum_{j=1}^{N} r_j\ s_{kj} + \sum_{j=1}^{N} s_{kj}^2 \tag{5.35}$$

The first terms may be ignored as it is independent on k, the second term of the expansion is the inner product of the observation vector r and signal vector s_k, and the third term is the energy (E_k) of the transmitted signal $s_k(t)$. So, a modified decision rule can be drawn as

Observation vector r lies in region Z_i, if

$$\sum_{j=1}^{N} r_j s_{kj} - E_k/2 \ \textit{is}\ \text{maximum for}\ k = i$$

Figure 5.5 shows the example of decision regions for $M = 4$ signals and $N = 2$ dimensions, assuming the signals are transmitted with equal energy, E, and equal probability.

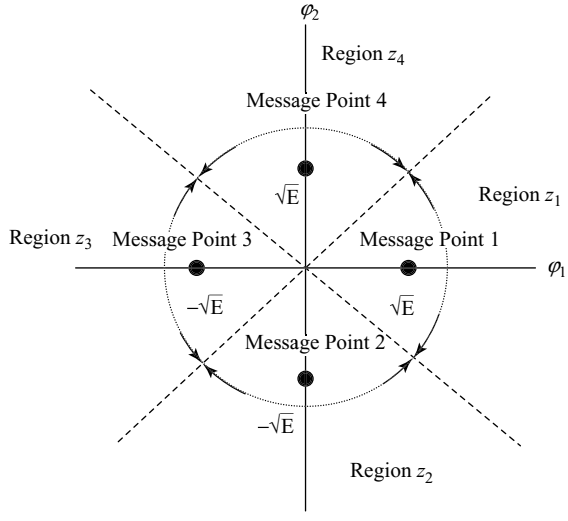


Fig. 5.5 *Example of decision regions for M = 4, N = 2*

5.8 OPTIMUM CORRELATION RECEIVER

Based on the previous analysis and discussions, we now discuss about the optimum correlation receiver for a AWGN channel and transmitted signals $s_1(t)$, $s_2(t)$, $\ldots s_M(t)$. The Optimum receiver has two parts:

1. The decoder part as illustrated in Fig. 5.6(a). It consists of M correlators or product integrators, along with N orthonormal basis functions $\varphi_1(t)$, $\varphi_2(t)$, $\varphi_3(t)$, $\ldots \varphi_N(t)$ that are locally generated. This bank of correlators operates on the received signal r(t) over $0 \le t \le T$, to produce the observation vector r.
2. The second part is the signal transmission decoder as shown in Fig. 5.6(b) which is based on maximum likelihood decoder. It operates on observation vector r to estimate m of the transmitted symbols m_i , $i = 1,2, \ldots M$ in such a way that the probability of errors become minimum.

There are N elements of observation vector r that are first multiplied by corresponding N elements of each M signal vectors $s_{1,}\, s_{2,}\, \ldots s_M$ and the resulting products are summed up in the accumulators to form the cor-

responding set of inner products $\{r^T s_k\}$, $k = 1,2,\ldots M$. The largest set is selected to estimate the transmitted message $\hat{m}$.

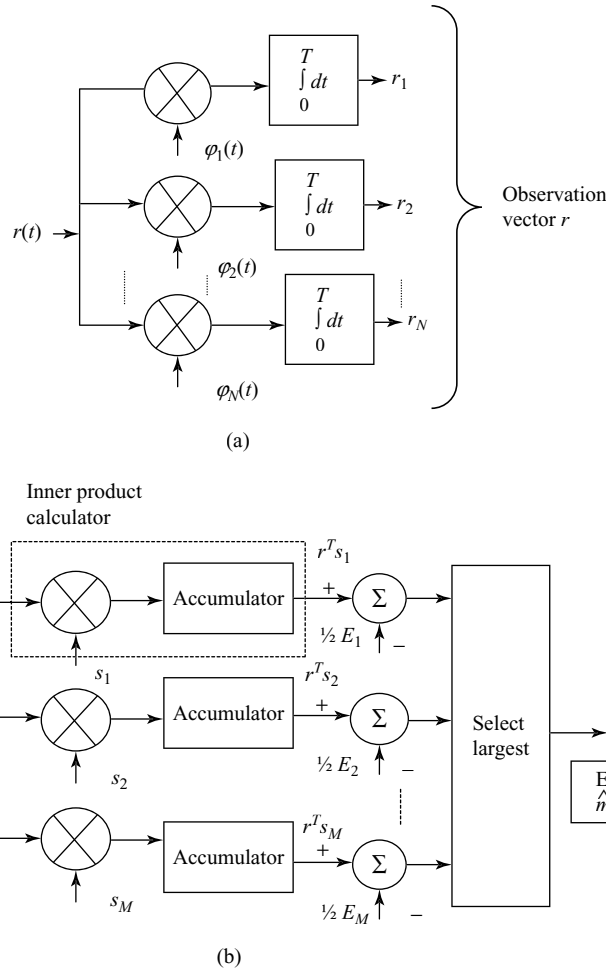


Fig. 5.6. (a) *Demodulator or detector,* (*b*) *Signal transmission decoder*

5.9 PROBABILITY OF ERROR

In a communication system, noise is always added with the transmitted signal and receiver may detect the signal wrongly. Considering a digital binary transmission system where symbols 1 and 0 are represented by positive and negative rectangular pulses of equal amplitude A and equal duration T_b (bit duration). The channel noise is AWGN with zero mean and power spectral density $N_0/2$. Within the signaling duration $0 \le t \le T_b$, the received signal is considered of the form,

$$r(t) = \begin{cases} +A + n(t) & \text{when symbol 1 is sent} \\ -A + n(t) & \text{when symbol 0 is sent} \end{cases} \tag{5.37}$$

Given the noisy signal $r(t)$, the receiver is required to make a decision in each signaling interval whether transmitted symbol is a 1 or a 0. Consider Fig. 5.7 as the receiving system for the binary transmitted signal.

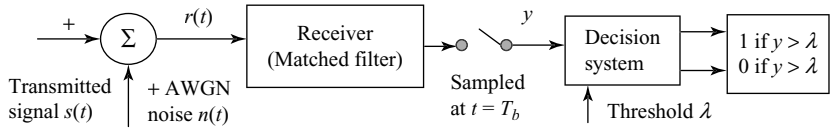


Fig. 5.7 *Receiver system for binary transmitted signal*

The filter is matched to the rectangular pulse of amplitude A and duration T_b exploiting the bit timing information available at the receiver. The matched filter output is random in nature because of the additive noise. If y is the sampled value at the end of signaling interval and compared with a predefined threshold λ in the decision system then symbol 1 is detected if y exceeds λ and decision goes in favor of symbol '1', '0' otherwise.

In the decision making process, there is always involvement of two types of error:

1. Symbol 1 is chosen when 0 was actually sent, we call this error as the Type I error.
2. Symbol 0 is chosen when 1 was actually sent; we call this error as the Type II error.

To determine the average probability of error, we consider these two situations separately.

Let p_{10} denote the conditional probability of error, given that symbol 0 was sent and p_{01} is the conditional probability of error, given that symbol 1 was sent. These probabilities are defined as follows:

$$\begin{aligned} p_{10} &= P(y > \lambda \mid \text{symbol 0 was sent}) \\ &= \int_{\lambda}^{\infty} f_Y(y \mid 0)\, dy \\ &= 1/\sqrt{\pi N_0 / T_b} \int_{\lambda}^{\infty} \exp\left(-\frac{(y+A)^2}{N_0 / T_b}\right) dy \end{aligned} \tag{5.38}$$

Similarly,

$$\begin{aligned} p_{01} &= P(y > \lambda \mid \text{symbol 1 was sent}) \\ &= \int_{-\infty}^{\lambda} f_Y(y \mid 1)\, dy \\ &= 1/\sqrt{\pi N_0 / T_b} \int_{-\infty}^{\lambda} \exp\left(-\frac{(y-A)^2}{N_0 / T_b}\right) dy \end{aligned} \tag{5.39}$$

Figure 5.8 shows the graphical representation of p_{10} and p_{01}.

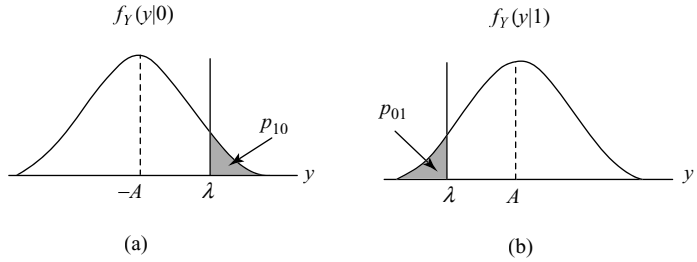


Fig. 5.8. *Graphical representation of probability density function of random variable Y at the matched filter output when (a) 0 is transmitted, (b) 1 is transmitted.*

Let us consider the complimentary error function erfc(u) which is closely related with Gaussian distribution as follows

$$\text{erfc}(u) = \frac{2}{\sqrt{\pi}} \int_{u}^{\infty} \exp(-z^2)\, dz \tag{5.40}$$

The upper bound of the complementary error function:

$$\text{erfc}(u) < \exp(-u^2) / \sqrt{\pi}\, u \tag{5.41}$$

In terms of error function, p_{10} can be expressed as

$$p_{10} = \frac{1}{\sqrt{\pi}} \int_{(A+\lambda)/\sqrt{N_0/T_b}}^{\infty} \exp(-z^2)\, dz = 1/2\ \text{erfc}(A + \lambda) / \sqrt{N_0/T_b} \tag{5.42}$$

where $z = (y + A) / \sqrt{N_0 / T_b}$

Similarly, error function p_{01} can be expressed as

$$p_{01} = \frac{1}{\sqrt{\pi}} \int_{(A-\lambda)/\sqrt{N_0/T_b}}^{\infty} \exp(-z^2)\, dz = 1/2\ \text{erfc}(A - \lambda) / \sqrt{N_0/T_b} \tag{5.43}$$

where $z = (A - y) / \sqrt{N_0 / T_b}$

The average probability of symbol error p_e can be expressed in terms of p_{10} and p_{01}. These two errors are mutually exclusive events. Let the a priori probabilities of transmitting symbols 0 and 1 are p_0 and p_1 respectively.

$$\begin{aligned} p_e &= p_0\ p_{10} + p_1\ p_{01} \\ &= p_0\ /2\ \text{erfc}(A + \lambda) / \sqrt{N_0/T_b}) + p_1\ /2\ \text{erfc}(A - \lambda) / \sqrt{N_0/T_b}) \end{aligned} \tag{5.44}$$

So, the average probability of error p_e is function of the threshold λ, which needs to be optimized to get minimum error probability p_e. When p_0 and p_1 are equi-probable, then $p_0 = p_1 = 1/2$. In that situation, the threshold is chosen at the middle between the pulse height $-A$ and $+A$ representing the two symbol 0 and 1as shown in Fig. 5.9.

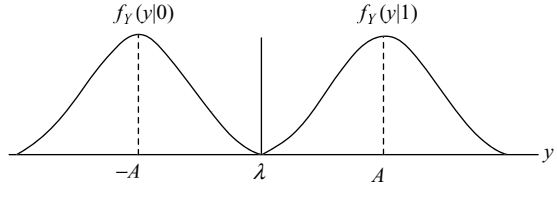


Fig. 5.9 *Threshold for binary symmetric channel*

For this special case of $p_{01} = p_{10}$, the channel is binary symmetric and the average probability of symbol error p_e becomes,

$$p_e = 1/2\ \text{erfc}(A / \sqrt{N_0 / T_b}) \tag{5.45}$$

The transmitted symbol energy per bit is

$$E_b = A^2 T_b \tag{5.46}$$

We finally formulate the p_e in terms of transmitted symbol energy per bit to the noise spectral density E_b / N_0 ratio solely.

$$p_e = 1/2 \text{ erfc } (\sqrt{E_b/N_0}) \tag{5.47}$$

With this background, we now express the probability of error for the correlation receiver illustrated in Fig. 5.6. Let that observation space Z is partitioned according to maximum likelihood decision rule into set of M regions $\{Z_i\}$, $i = 1,2, \ldots M$. When symbol m_i or signal vector s_i is transmitted, let the observation vector r is received. Error will occur when the observation vector r does not lie within region Z_i associated with the message point represented by s_i. Averaging of all possible transmission, the average probability of error p_e is defined as,

$$\begin{aligned} p_e &= \sum_{i=1}^{M} p_i \; P(r \text{ does not lie in } Z_i \mid m_i \text{ is sent}) \\ &= 1/M \sum_{i=1}^{M} P(r \text{ does not lie in } Z_i \mid m_i \text{ is sent}) \\ &= 1 - 1/M \sum_{i=1}^{M} P(r \text{ lies in } Z_i \mid m_i \text{ is sent}) \end{aligned} \tag{5.48}$$

In terms of likelihood function,

$$p_e = 1 - 1/\, M \sum_{i=1}^{M} \int_{Zi} f_r(r \mid m_i)\, dr \tag{5.49}$$

5.10 TYPES OF DIGITAL MODULATION TECHNIQUES

1. **Coherent digital modulation techniques:** Coherent digital modulation techniques use coherent detection at the receiver using locally carrier generated at the receiver with phase locked at the transmitting carrier. Thus, correlating received noisy signal and locally generated carrier do the detection, which is a synchronous detection process.
2. **Non-coherent digital modulation techniques:** Non-coherent digital modulation techniques do not use receiver carrier to be phase locked with the transmitter carrier. Thus, the advantage of this process is the simplicity for hardware implementation. But, the disadvantage is that error probability increases.

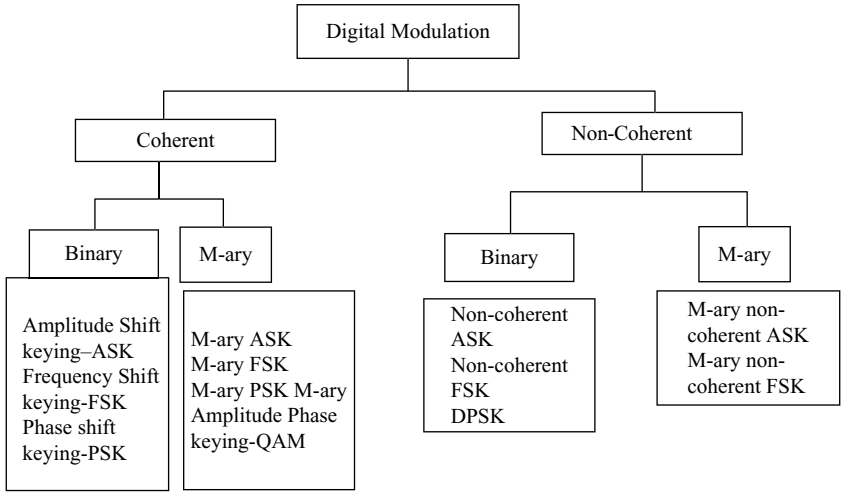


Fig. 5. 10 *Different types of digital modulations*

Based on these two techniques Fig. 5.10 shows the different types of digital modulation techniques.

In an M-ary signaling scheme, we may send any one of M possible signals $s_1(t), s_2(t), \ldots s_M(t)$, during each signaling interval of duration T. The number of possible signals $M = 2^n$, n is an integer. The symbol duration $T = nT_b$, where T_b is the bit duration. In baseband data transmission, these signals are generated in different form by changing amplitude, phase and frequency of a sinusoidal carrier in M discrete steps. In this way, M-ary ASK, M-ary PSK and M-ary FSK digital modulation schemes are obtained. A special form of modulation namely M-ary quadrature-amplitude modulation (QAM) is obtained by a hybrid technique. M-ary signaling is used when there is a requirement of conserving bandwidth at the cost of higher power. When the bandwidth of the channel is less than the required value, M-ary modulation is used for maximum bandwidth efficiency.

A major goal of digital baseband transmission is to design optimum receiver to minimize the average error probability of symbol error in presence of AWGN.

5.11 COHERENT BINARY MODULATION TECHNIQUES

In the baseband digital systems, signal transmission is done without shifting the frequency of the signal. Baseband signal transmission has finite power at low frequencies and suitable for transmission over a pair of wires, coaxial cables or optical fibres. But this baseband signal cannot be transmitted over radio link or satellite because this would necessitate impractically large antenna to transmit low frequency signal efficiently. That is why low frequency signal is shifted to high frequency at first through some modulation scheme. The binary digital modulation has three basic forms: Amplitude Shift Keying (ASK), Frequency Shift Keying (FSK) and Phase Shift Keying (PSK). In the following sections, we shall describe the different coherent binary modulation techniques.

1. **Amplitude Shift Keying, ASK or On-Off Keying (OOK):** An unmodulated carrier of the form $\sin \omega_c t$, an on-off baseband signal $m(t)$ also called modulating signal, the carrier signal amplitude is varied in accordance with the baseband signal present. The modulated signal is represented as $m(t)$ $\sin \omega_c t$ which is still on-off signal. This way of transmitting binary data is known as OOK or ASK.

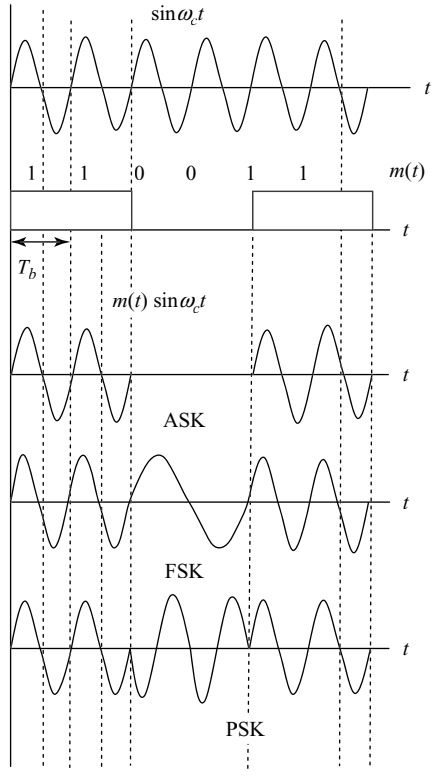


Fig. 5. 11 *Digital modulation waveforms*

Let for ASK wave form the signal $s(t) = 0$, when no transmission happens, i.e., the symbol 0, and $s(t) = \sqrt{2P_s} \sin \omega_c t$ for transmitting the symbol 1. $s(t)$ contains the complete cycles of carrier frequency f_c.Where, $P_s = E_b/T_b$, unit bit signal energy, T_b is the bit duration.

$$s(t) = \sqrt{P_s T_b}\ \varphi_1(t) = \sqrt{E_b}\,\varphi_1(t) \tag{5.50}$$

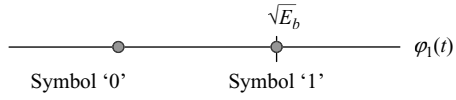


Fig. 5.12 *Signal space representation for ASK*

So, the distance between the two message points is $d = \sqrt{E_b}$.

2. **Frequency Shift Keying FSK:** An un-modulated carrier is sin $\omega_c t$, for the modulating signal (base band) $m(t)$, the carrier signal frequency is varied in accordance with the baseband signal. The modulated signal is represented as $m(t)\sin \omega_c t$ which has two frequency componets. Baseband signal 1 (high) corresponds frequency f_1, and 0 (low) corresponds frequency f_2 in the modulated signal. This way of transmitting binary data is known as FSK.
3. **Phase Shift Keying PSK:** For PSK there is a change of phase when baseband signal goes from high to low or low to high. This is illustrated in Fig. 5.11.

5.12 COHERENT BINARY PHASE SHIFT KEYING: BPSK

In a coherent binary PSK system, symbols 1 and 0 are represented by the pair of signals $s_1(t)$ and $s_2\ (t)$ respectively as,

$$s_1(t) = \sqrt{2E_b/T_b}\ \cos(2\pi f_c\ t) \tag{5.51}$$

and

$$s_2(t) = -\sqrt{2E_b/T_b}\ \cos(2\pi f_c\ t) \tag{5.52}$$

$$\text{for} \quad 0 \le t \le T_b$$

E_b is the transmitted signal energy /bit and T_b is the bit duration. Each transmitted bit contains number of cycles of the carrier wave. The carrier frequency $f_c = n$ x $1/T_b$, where n is an integer, then the integral number of cycles of the carrier wave can be accommodated during transmission. A pair of sinusoidal waves differs only by relative phase shift of 180^0 as shown in Fig. 5.13.
So, for BPSK only one basis function with unit energy is defined:

$\varphi_1(t) = \sqrt{2/T_b}\ \cos(2\pi f_c t)$ for $0 \le t \le T_b$ and thus $s_1\ (t)$ and $s_2\ (t)$ can be defined as,

$$s_1(t) = \sqrt{E_b}\ \varphi_1(t) \text{ and} \tag{5.53}$$

$$s_2(t) = -\sqrt{E_b}\ \varphi_1(t) \quad \text{for} \quad 0 \le t \le T_b \tag{5.54}$$

The coherent BPSK system has one dimensional signal space and two message points. The coordinates of the message points are:

$$s_{11} = \int_0^{T_b} s_1(t)\ \varphi_1(t)\ dt = +\sqrt{E_b} \tag{5.55}$$

$$s_{12} = \int_0^{T_b} s_2(t)\ \varphi_1(t)\ dt = -\sqrt{E_b} \tag{5.56}$$

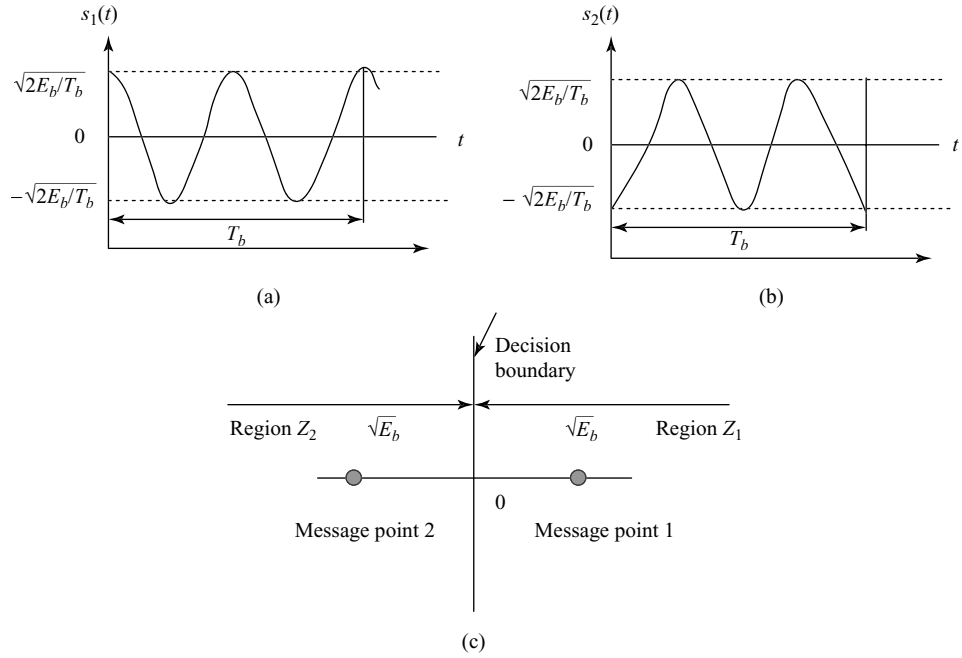


Fig. 5.13 *Signal space representation for (a) $s_1(t)$, (b) $s_2(t)$ of transmitted signal with N = 2, (c) Message points for BPSK*

Thus for BPSK two message points lie at $s_{11} = \sqrt{E_b}$ and $s_{12} = -\sqrt{E_b}$ corresponding to the signals $s_1(t)$ (for 1 bit) and $s_2(t)$ (for 0 bit) respectively.

5.12.1 Geometrical Representation of BPSK Signal

The two signals for BPSK is combined as $s(t) = \pm \sqrt{E_b}\ \varphi_1(t)$ (5.57)

So, on the single axis of $\varphi_1(t)$, there will be two message points located at $+\sqrt{E_b}$ and $-\sqrt{E_b}$ as shown in Fig. 5.13. At the receiver, the point at $+\sqrt{E_b}$ on $\varphi_1(t)$ represents symbol 1 and the point at $-\sqrt{E_b}$ represents symbol 0. The separation between these two points is $d = 2\sqrt{E_b}$. As this distance increases, isolation between the symbols would be better, thus reduces the error probability.

5.13 ERROR PROBABILITY OF BPSK

The signal space of Fig. 5.13 is divided into two regions: the set of points closest to message point 1 at $+\sqrt{E_b}$ and the set of points closest to message point 0 at $-\sqrt{E_b}$. The two decision regions are marked as Z_1 and Z_2 respectively. The decision rule is to decide that signal $s_1(t)$ (symbol 1) is transmitted if the received signal falls within Z_1 and that for $s_2(t)$ (symbol 0) if the received signal falls within region Z_2. To make this decision two kinds of erroneous decision may occur as discussed in earlier section. Let the decision region associated with symbol 1 or signal $s_1(t)$ is described as,

$Z_1 : 0 < r_1 < \infty$ where the observed signal element r_1 is related to the received signal $r(t)$.

$$r_1 = \int_0^{T_b} r(t)\ \varphi_1(t)\ dt \tag{5.58}$$

The conditional probability density function of random variable R_1, given that symbol 0 (signal $s_2(t)$) is transmitted is defined as,

$$f_{RI}(r_1 \mid 0) = \frac{1}{\sqrt{\pi N_0}} \exp\left(-1/N_0 (r_1 - s_{21})^2\right)$$
$$= \frac{1}{\sqrt{\pi N_0}} \exp\left(-1/N_0 (r_1 + \sqrt{E_b})^2\right) \quad (5.59)$$

The conditional probability p_{10} of the receiver deciding in favor of the symbol 1, given that symbol 0 was transmitted is therefore given as

$$p_{10} = \int_0^\infty f_{R_1}(r_1 \mid 0)\, dr_1$$
$$= \frac{1}{\sqrt{\pi N_0}} \int_0^\infty \exp\left(-1/N_0\ (r_1 + \sqrt{E_b})^2\right) dr_1 \quad (5.60)$$

$$Let\ z = (1/\sqrt{N_0})\ (r_1 + \sqrt{E_b})$$
$$p_{10} = \frac{1}{\pi} \int_{\sqrt{E_b/N_0}}^\infty \exp(-z^2)\, dz$$
$$= 1/2\ \text{erfc}\left[\sqrt{E_b / N_0}\right] \quad (5.61)$$

Similarly, it can be proved that the conditional probability p_{01} of the receiver deciding in favor of the symbol 0, given that symbol 1 was transmitted is given also as

$$p_{01} = 1/2\ \text{erfc}\left[\sqrt{E_b / N_0}\right]$$

So the average probability of detecting symbol error or bit error rate for coherent BPSK is

$$p_e = 1/2\ \text{erfc}\left[\sqrt{E_b / N_0}\right] \quad (5.62)$$

As we increase the energy per bit E_b for a specified noise spectral density, the message points $\sqrt{E_b}$ and $-\sqrt{E_b}$ corresponding to 1 and 0 move further apart. So, probability of error detection will be reduced. At this stage it may be inferred that erroneous detection can be avoided by increasing the power of the transmitter, however, it is not always the only remedy from real engineering point of view.

5.14 GENERATION AND DETECTION OF BPSK SIGNAL

To generate the BPSK signal, the binary sequence is encoded first into polar non-return to zero (NRZ) signals. The binary sequence in polar form represents the symbol 1 and 0 by constant amplitude levels of $\sqrt{E_b}$ and $-\sqrt{E_b}$ respectively.

A sinusoidal carrier $\varphi 1(t) = \sqrt{2/T_b}\cos(2\pi f_c t)$ is used to a product modulator to generate the BPSK signals represented by Eqs. (5.51–5.52), where $f_c = n/T_b$. For synchronization, a common master clock is used to extract timing information both for carrier and pulses. The signal is then transmitted over noisy channel where AWGN is added to get received signal $r(t)$ which is applied at the correlator input for the detection circuit. A locally generated coherent signal $\varphi_1(t)$ is supplied at the correlator and multiplied with the $r(t)$, integrated over the signal duration T_b to get some output value r_1 that would be compared with the reference signal called threshold (= 0 volts) in the decision devices to detect the symbol '0' or '1'. If r_1 greater than threshold then symbol 1 is detected, if r_1 is less than the threshold then symbol '0' is detected. If r_1 is exactly same to threshold then receiver makes a random guess in favor of 1 or 0. Fig. 5.14 is the block diagram for generation and detection of BPSK signal.

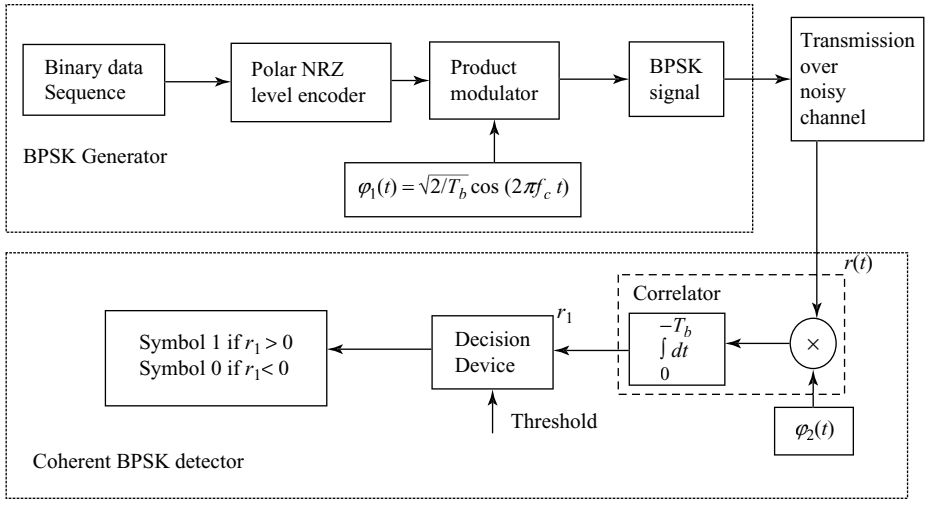


Fig. 5.14 *Generation and detection of coherent BPSK*

5.14.1 Testbed Implementation of BPSK

Multiplication is a basic function of modulation. In our testbed development, we have used the idea of multiplication with some other operation such as level shifting and sum. The ICL8013 is a four quadrant analog multiplier whose output is proportional to the algebraic product of two input signals (analog or digital). Feedback around an internal OPAMP provides level shifting. The ICL8013 has features such as high accuracy, wide bandwidth, and increased versatility, which made it suitable for all multiplier applications in control and instrumentation systems. Its analog multiplication, nature makes it an ideal basic component for implementing modulation and demodulation circuits (analog or digital scheme) in laboratory environment communication system. Functional Block Diagram and Pin diagram of Analog Multiplier is shown in Fig. 5.15. Figure 5.16 is the circuit for the generation of BPSK modulation, where the X_{IN} is the sinusoidal carrier input and Y_{IN} is the bipolar data input. Figure 5.17 is the practical testbed implementation of BPSK modulator, where data is generated by 4 bit decade counter (74L*S*390) and a sinusoidal carrier as taken from function generator is applied to X_{IN} of the multiplier.

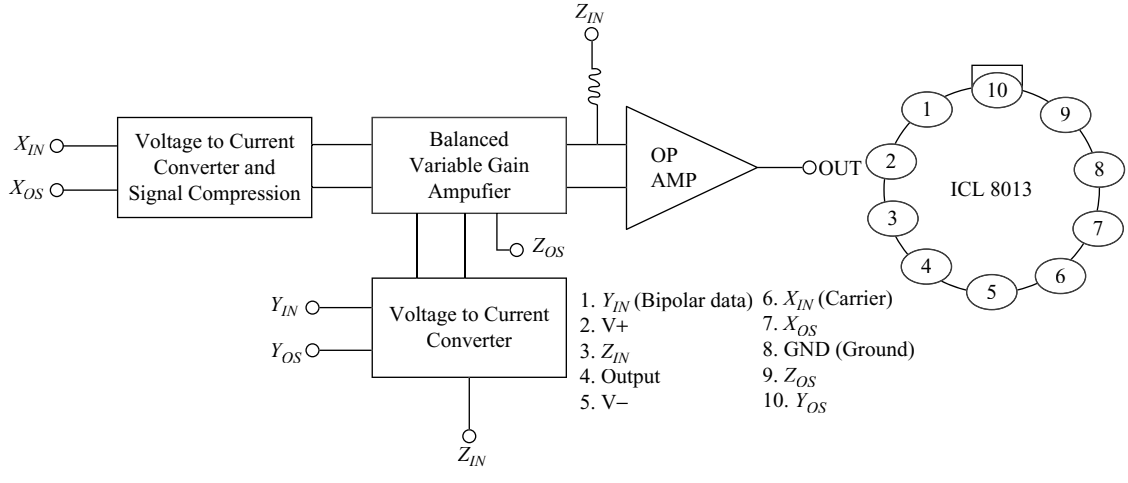


Fig. 5.15 *Functional and pin diagram of analog multiplier ICL 8013*

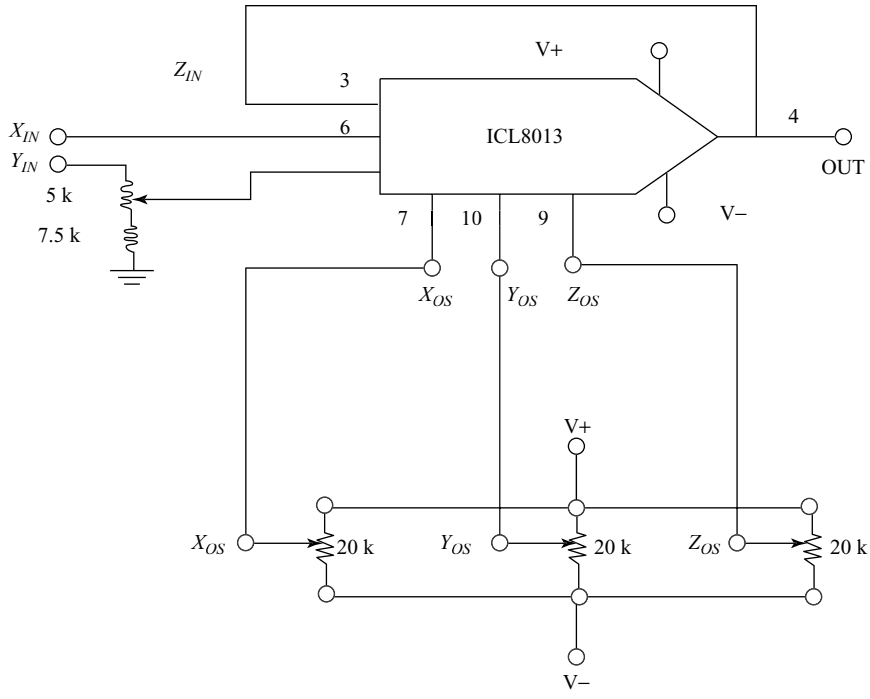


Fig. 5.16 *BPSK modulator circuit diagram using ICL8013*

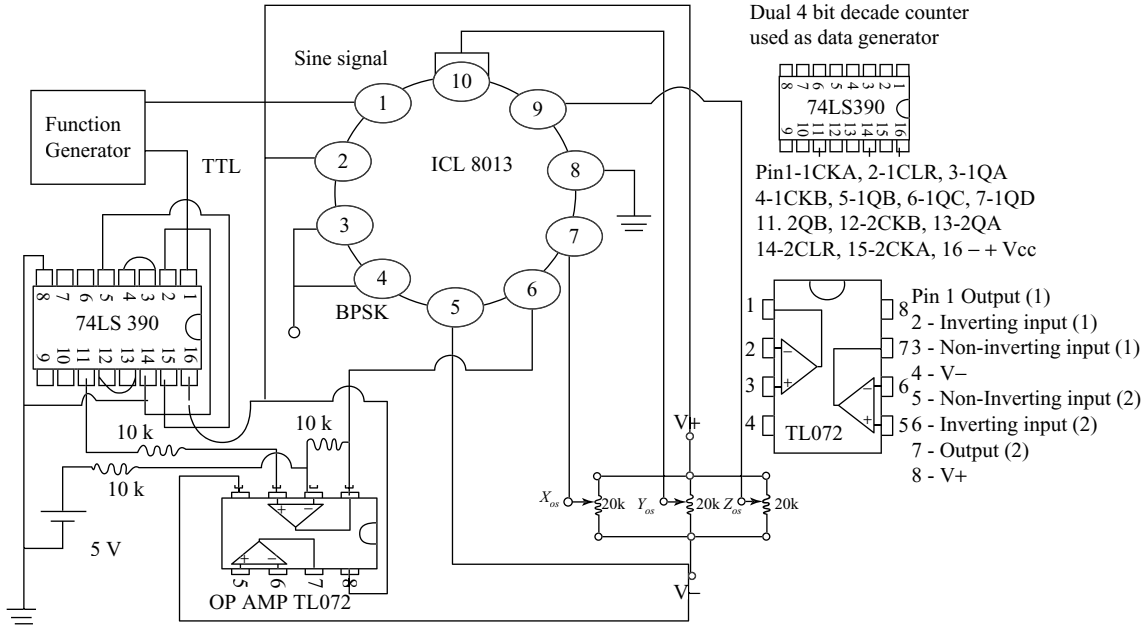


Fig. 5. 17 *Practical testbed of BPSK modulator*

The Op-Amp TL072 is used to convert the unipolar data bits to a bipolar signal as required by the analog multiplier ICL 8013 for proper multiplication to produce BPSK modulated signal. The principle of the unipolar to bipolar conversion circuit in our system is the level shifting of the data stream and amplification

by a factor of 2. The Op Amp is in non-inverting configuration as the signal is applied at the non-inverting terminal. So, the gain of the signal is $(1 + (R_f/R_I)) = 2$, as R_f = 10k and R_I = 10k and to shift the level of the signal, a positive voltage of 5V is applied at inverting terminal to subtract from the amplified signal. The range of resultant signal becomes +5V and −5V.

Figure 5.18 is the experimentally observed output of the BPSK modulator with a data stram of 01010010. The waveforms from the BPSK modulator have been observed through a 4-channel digital storage ocillascope. The channel 1 in the Figure is the Data and the channel 3 is BPSK modulated signal.

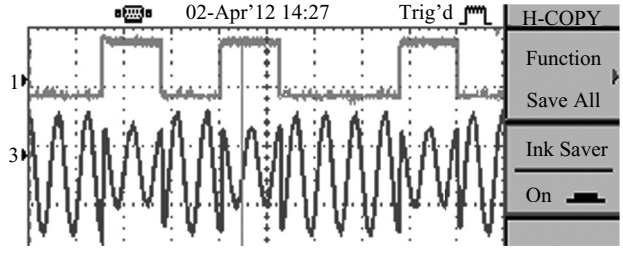


Fig. 5. 18 *Testbed output for BPSK modulation*

5.15 POWER SPECTRUM OF BPSK

In the BPSK waveform, there are rectangular pulses of amplitude $\pm\sqrt{2E_b/T_b}$ over the duration $0 \le t \le T_b$ depending on whether we have symbol 1 or 0. So, the symbol shaping function $g(t)$ of the BPSK signal can be defined as,

$$g(t) = \begin{cases} \sqrt{2\,E_b/T_b} & 0 \le t \le T_b \\ 0 & \text{Otherwise} \end{cases} \tag{5.63}$$

The input binary waves are random in nature. If we consider that 1 and 0 transmissions are equally likely and the symbols transmitted at different times are statistically independent, then over large number of symbols transmission, the power spectral density can be averaged over the duration of the transmission. The energy spectral density of the base signal is found by getting the Fourier transform of the g(t) which has a sinc function distribution. The energy spectral density is the squared magnitude of the signal's Fourier transform. The Fourier transform of g(t) is

$$G(f) = \sqrt{2E_b/T_b}\; T_b\; \sin c\,(\pi f T_b) = A\,T_b \sin c\,(\pi f T_b), A = \sqrt{2E_b/T_b}$$
$$G(\omega) = AT_b \sin c\,(\omega\, T_b/2) \tag{5.64}$$

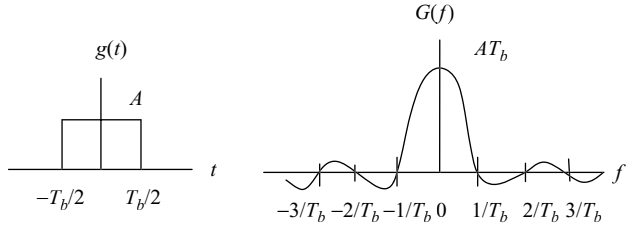


Fig. 5.19 *Plot for pulse shaping function and it's Fourier transform*

$$\text{The energy spectral density (ESD)} = |\,G(\omega)|^2 = A^2T_b^{\,2} \sin c^2(\omega\, T_b/2) \tag{5.65}$$
$$= 2E_bT_b \sin c^2(\omega\, T_b/2)$$

The power spectral density (PSD) $S(\omega) = \text{ESD}/T_b = 2E_b \sin c^2(\omega T_b/2)$ (5.66)

The BPSK signal is generated by modulating the baseband signal $g(t)$ with the carrier $\cos(2\pi f_c t)$ to produce $g(t)\cos(2\pi f_c t)$ and thus spectral components are frequency translated to $f \pm f_c$ and the magnitude becomes half.

$$g(t)\cos(2\pi f_c t) \overset{FT}{\Leftrightarrow} 1/2\,[G(\omega+\omega_c)+G(\omega-\omega_c)] \tag{5.67}$$

Figure 5.20 is the plot for PSD $S(\omega)$ for baseband signal g(t) and BPSK signal $S_{\text{BPSK}}(\omega)$.

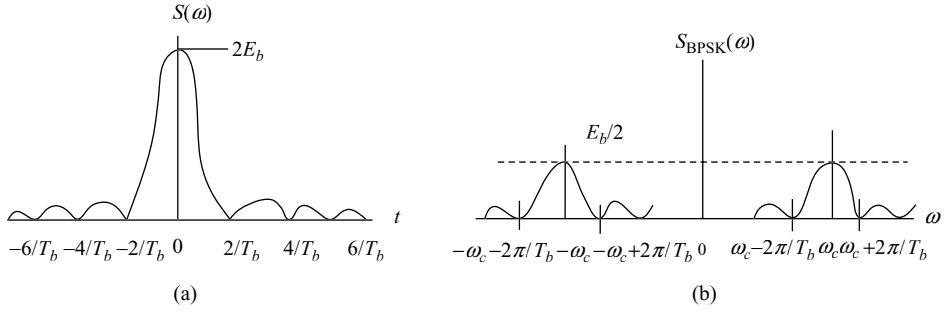


Fig. 5.20 *Power Spectral Density of* (*a*) *NRZ baseband signal* $g(t)$, *and* (*b*) *BPSK signal*

The minimum bandwidth of BPSK signal is equal to twice of the highest frequency contained in the baseband signal.

5.16 CONCEPT OF M-ARY COMMUNICATION

So far we have discussed about the binary transmission where two symbols are used only. In contrary for M –ary case, the total number of symbols used is M. The information carried by each M-ary symbol is $\log_2 M$ binary digits. In order to obtain a given performance level, M-ary communication needs more transmission power (multi-amplitude, multiphase case) and more bandwidth (multitone case).

In M-ary signaling, the source symbol can take M possible number of integers and the amplitude or phase of the carrier takes on one of M possible values.

Thus $S \in \{\pm 1, \pm 2, \ldots\ldots \pm (M-1)\}$, there are M possible waveforms; during a symbol interval T_s one of the waveforms is selected for transmission. The M-ary communication may be multi-amplitude (MASK) where M symbols are transmitted by M-pulses $\pm g(t), \pm 3g(t), \pm 5g(t), \ldots\ldots \pm (M-1)g(t)$. In case of multi-phase (MPSK) signaling M pulses with phases of successive pulse of $2\pi/M$ radians would be transmitted.

5.16.1 M-Ary Phase Shift Keying (MPSK)

The phase of the carrier takes on one of M possible values such as $\theta_i = 2\pi(i-1)/M$, where $i = 1,2,3,\ldots,M$. For each signaling duration T_s, one of M possible waveform is represented as

$$s_i(t) = \sqrt{2E_s/T_s}\cos\left[2\pi f_c\, t + 2\pi(i-1)/M\right] over\ 0 \le t \le T_s \tag{5.68}$$
$$i = 1, 2, \ldots..M$$

where, E_s is the energy of the symbol $= \int_{-\infty}^{+\infty} s_i^2(t)\,dt$

The carrier frequency $f_c = n/T_s$, for any integer n. Each of $s_i(t)$ may be expanded in terms of two basis functions $\varphi_1(t)$ and $\varphi_2(t)$.

$$\varphi_1(t) = \sqrt{2/T_s}\cos 2\pi f_c\, t \quad \text{and} \quad \varphi_2(t) = \sqrt{2/T_s}\sin 2\pi f_c\, t \quad \text{for } 0 \le t \le T_s$$
$$i = 1,2,\ldots M$$

Thus the signal constellation of MPSK is two-dimensional.

If E_b and T_b represent the bit energy and bit duration, then the symbol energy $E_s = nE_b$ and symbol duration $T_s = nT_b$. Unlike the binary phase shift keying the signal constellation for M-ary PSK is two-dimensional. The M-message points lies on the circle of radius $\sqrt{E_s}$ and center at the origin as shown in Fig. 5.21.

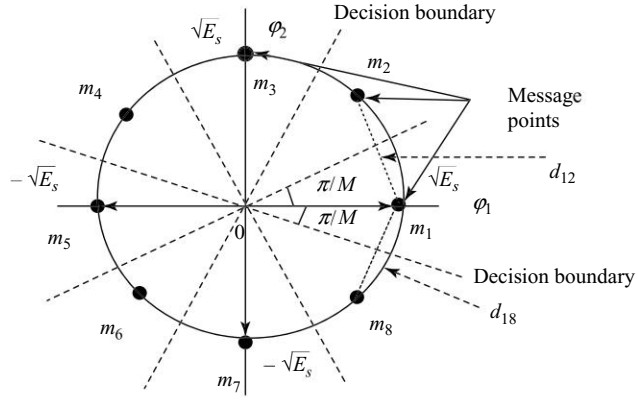


Fig. 5.21 *Signal space diagram for M-ary PSK*

5.16.2 Average Probability of Symbol Error for Coherent M-ary PSK

Suppose that the transmitted signal corresponds to the message point m_1 as shown in Fig. 5.17. The coordinate of m_1 on $\varphi_1(t)$ and $\varphi_2(t)$ axes are $\sqrt{E_s}$ and 0 respectively. We consider that E_s/N_0 ratio is large enough to detect the message point m_1 on the decision boundary. We know that the error probability depends on the Euclidean distance between the two message points on the signal space diagram. In Fig. 5.21, we consider M = 8, as there are 8 message points and the Euclidean distance between two message points is,

$$d_{12} = d_{18} = 2\sqrt{E_s}\sin(\pi/M) \tag{5.69}$$

So, the average error probability of symbol error for coherent MPSK is

$$P_e = \text{erfc}\left[\sqrt{E_s/N_0}\sin(\pi/M)\right] \tag{5.70}$$

We assume $M >= 4$.

5.16.3 Power Spectra of MPSK

For M-ary PSK, the symbol duration $T_s = T_b \log_2 M$, where T_b is the bit duration. As obtained in the BPSK case, the PSD is

$$S(\omega) = 2E_b \sin c^2(\omega T_b/2)$$

Substituting, E_b by E_s and T_b by T_s, the PSD for MPSK becomes

$$S_{MPSK}(\omega) = 2E_s \sin c^2(\omega T_s/2) = 2E_b \log_2 M \sin c^2(T_b f \log_2 M) \tag{5.71}$$

The normalized power spectra of MPSK for different M with respect to normalized frequency fT_b is shown in Fig. 5.22.

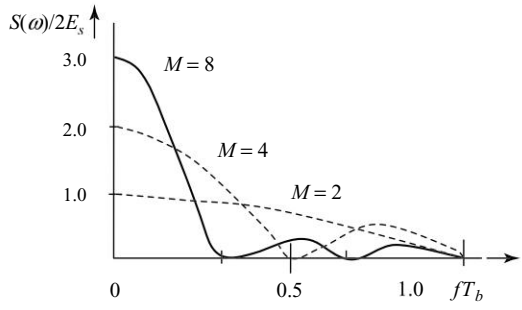


Fig. 5.22 *Normalized power spectra for MPSK*

The main lobe in the power spectra simply provides the information for essential bandwidth of the MPSK signals that contains most of the power. The channel pass band bandwidth for MPSK is $B_{\text{MPSK}} = 2/T_s$, where T_s is the symbol duration. Replacing T_s in terms of T_b and expressing $R_b = 1/T_b$, the bit rate, the channel bandwidth can be redefined as

$$B_{\text{MPSK}} = 2/T_b \log_2 M = 2R_b / \log_2 M \tag{5.72}$$

The channel efficiency ρ_{ch} is defined as = Bit rate/Bandwidth

$$\rho_{ch} = R_b / B_{\text{MPSK}} = \log_2 M / 2 \tag{5.73}$$

As M, the number of possible states increases, the channel efficiency will increase, but probability of error will also increase. As a trade-off parameters the E_b/N_0 ratio needs to be adjusted properly (higher ratio) as M-increases.

5.17 QUADRATURE PHASE SHIFT KEYING: QPSK

QPSK is the special case of MPSK with $M = 4$, symbols can take four possible values (10, 00, 01, 11) each corresponds to 2 bits (dibits).

For QPSK signal, the phase of the carrier can take only four equal spaced values, such as, $\pi/4$, $3\pi/4$, $5\pi/4$ and $7\pi/4$. The transmitted signal is given by

$$s_i(t) = \begin{cases} \sqrt{2\,E_s / T_s}\cos\left[2\pi f_c t + (2i-1)\,\pi/4\right], & \text{over } 0 \le t \le T_s \\ 0, & \text{Otherwise} \end{cases} \tag{5.74}$$

$i = 1, 2, 3$ and 4

Thus,

$$s_i(t) = \sqrt{2E_s/T_s}\cos\left[(2i-1)\,\pi/4\right]\cos 2\pi f_c t - \sqrt{2E_s/T_s}\sin\left[(2i-1)\,\pi/4\right]\sin 2\pi f_c t \tag{5.75}$$

In terms of basis functions,

$$\begin{aligned} s_i(t) &= \sqrt{E_s}\cos[(2i-1)\,\pi/4]\,\varphi_1(t) - \sqrt{E_s}\sin[(2i-1)\,\pi/4]\,\varphi_2(t) \\ &= s_{i1}\varphi_1(t) + s_{i2}\varphi_2(t) \end{aligned} \tag{5.76}$$

where

$$\varphi_1(t) = \sqrt{2/T_s}\cos 2\pi f_c t \qquad \varphi_2(t) = \sqrt{2/T_s}\sin 2\pi f_c t \qquad \text{for } 0 \le t \le T_s$$

So, in the two-dimensional signal space defined by the two-orthonormal sets $\{\varphi_1(t), \varphi_2(t)\}$, there are four message points for QPSK signal that can be defined by signal vector as

$$s_i = \begin{bmatrix} \sqrt{E_s} & \cos\left[(2i-1)\,\pi/4\right]\varphi_1(t) \\ -\sqrt{E_s} & \sin\left[(2i-1)\,\pi/4\right]\varphi_2(t) \end{bmatrix} \tag{5.77}$$

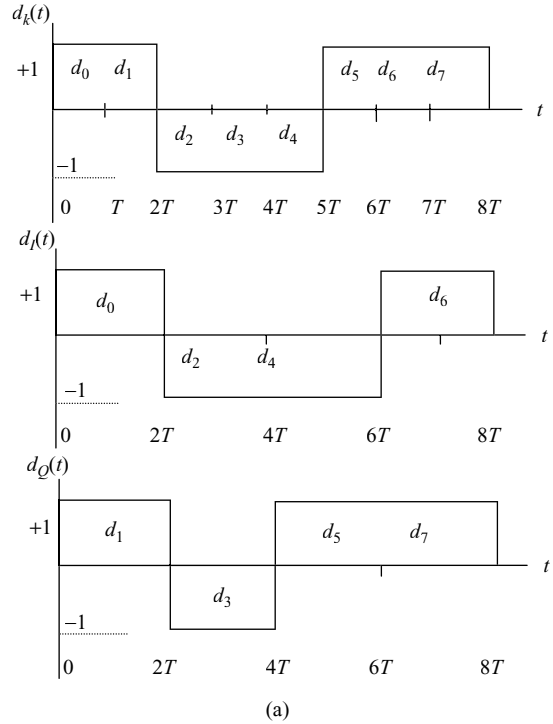
$i = 1, 2, 3$ and 4

Table 5.1 shows the elements of signal vectors s_{i1} and s_{i2}. Figure 5.23 shows the generation of QPSK signal from the input binary sequence considering Table 5.1 and Eq. (5.76), $d_k = \{0, 1\}$ arrives at the modulator input at a rate $1/T$ bits/sec and is separated into two data streams $d_I(t)$ and $d_Q(t)$ containing odd and even bits respectively.

Table 5.1 *Signal space characterization of QPSK*

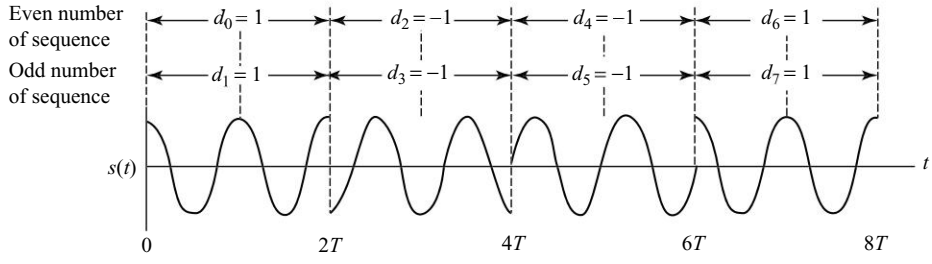
Input Dibit	Phase of QPSK signal in radians	Coordinates of message points s_{i1}	s_{i2}
10	$\pi/4$	$+\sqrt{E_s/2}$	$-\sqrt{E_s/2}$
00	$3\pi/4$	$-\sqrt{E_s/2}$	$-\sqrt{E_s/2}$
01	$5\pi/4$	$-\sqrt{E_s/2}$	$+\sqrt{E_s/2}$
11	$7\pi/4$	$+\sqrt{E_s/2}$	$+\sqrt{E_s/2}$

Each of the four possible phases of carriers represents two bits of data. Thus there are two bits per symbol. Since the symbol rate for QPSK is half the bit rate, twice as much data can be carried in the same amount of channel bandwidth as compared to BPSK. This is possible because the two signals I(inphase) and Q(Quadrature) are orthogonal to each other and can be transmitted without interfering with each other.



(a)

Fig. 5.23 *Generation of QPSK signal* (*a*) *The input binary bit stream* { d_k}, *odd and even bit sequences,* (*b*) *QPSK signal showing phase change.* (*continued*)



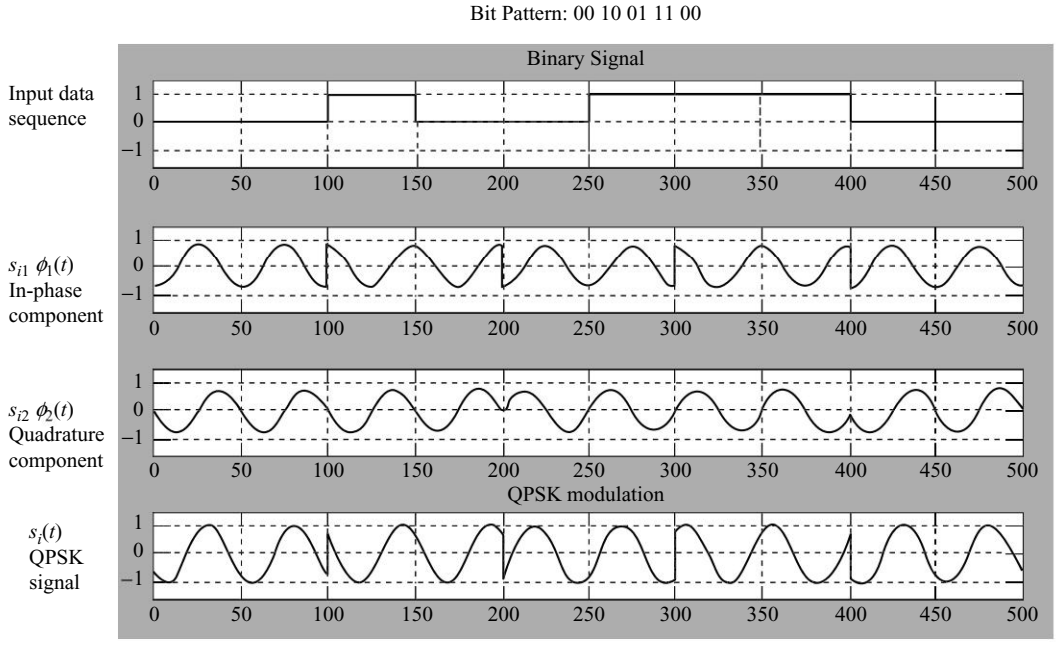
(b)

Fig. 5.23 (*continued*)

In QPSK the carrier phase can change only once every $2T$ secs. If from one T interval to the next one, neither bit stream changes sign, the carrier phase remains unchanged. If one component $d_I(t)$ or $d_Q(t)$ changes sign, a phase change of $\pi/2$ occurs (Example: when the binary sequence switches from dibit 10 to dibit 00).

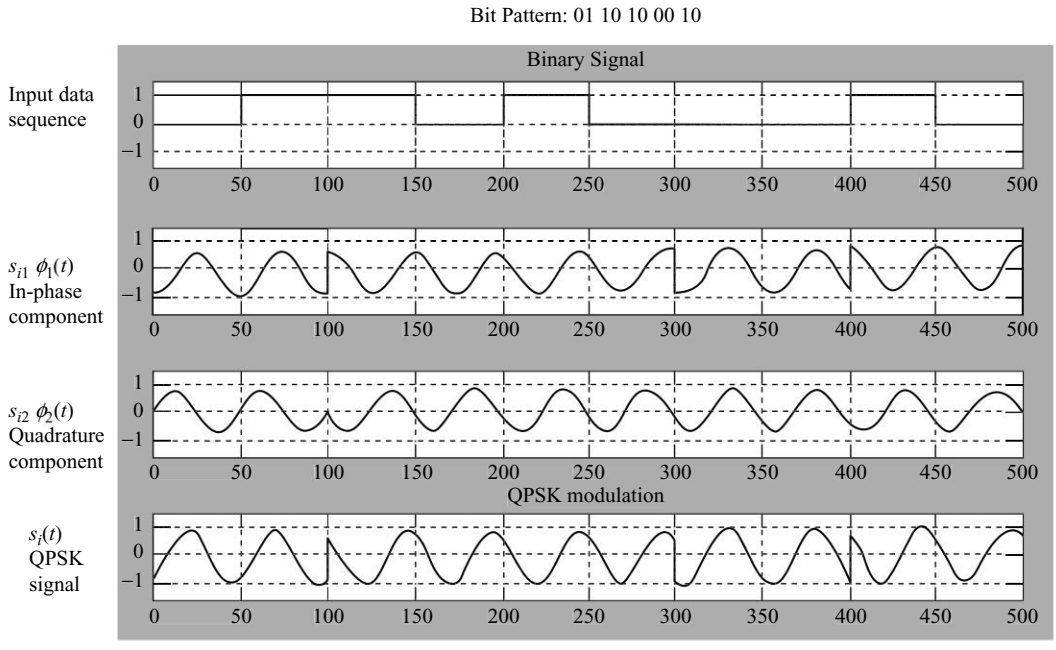
However, if both components change sign then a phase shift of π occurs (Example: when the binary sequence switches from dibit 01 to dibit 10).

The generation of QPSK waveforms can best be explained by the examples given in Fig. 5.24. Let the input binary data sequence in 5.24(a) be 0010011100. In QPSK modulation schemes dibits are transmitted and the corresponding phase of the QPSK signal undergoes a change of $\pi/2$ for successive change in dibits as shown in Table 5.1. The coordinates of the message points also exhibit no change/change in sign, according to the transmitted dibits. The signal space diagram for QPSK modulation involves two orthonormal basis functions; one is a cosine waveform $\varphi_1(t)$ while the other being a sine waveform $\varphi_2(t)$. The decomposition of the input data sequence into odd-numbered sequence or in-phase component and even numbered sequence or quadratute component along with the polarity of the co-efficients s_{i1} and s_{i2} are shown in Fig. 5.25. The first data bit is considered as the odd one followed by even bit and so on.

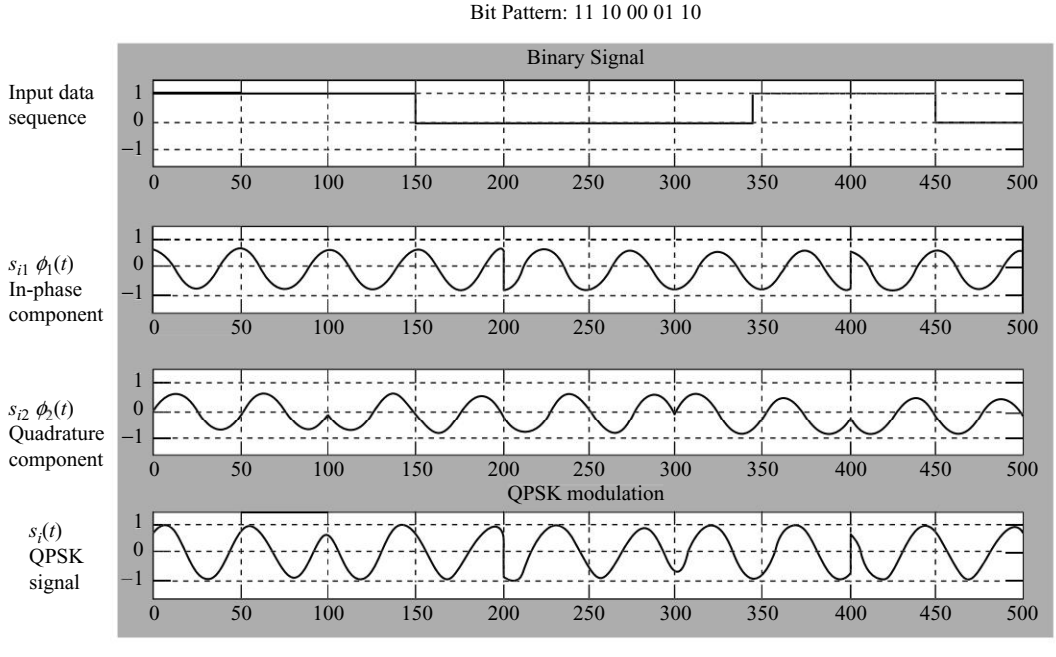


(a)

Fig. 5.24 *Examples of generating QPSK signals* (*continued*)



(b)



(c)

Fig. 5.24 (*continued*)

Now as per Table 5.1, the signs of the cosine and sine components of the QPSK waveform will vary in accordance with the variation in polarities of the coefficients s_{i1} and $s_{i2.}$ Table 5.2 represents the nature of the constituent waveform for the input data stream 0010011100 according to Fig. 5.25.

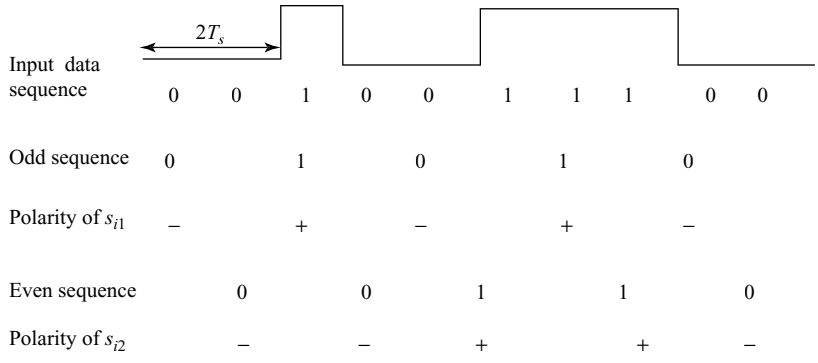


Fig. 5.25 *Odd-even sequences for QPSK signal showing the polarity*

Table 5.2 *QPSK waveforms corresponding to dibits*

Input dibit	Nature of $s_{i1}\,\varphi_1(t)$	Nature of $s_{i2}\,\varphi_2(t)$
00	Inverted Cosine	Inverted Sine
10	Cosine	Inverted Sine
01	Inverted Cosine	Sine
11	Cosine	Sine
00	Inverted Cosine	Inverted Sine

Thus the nature of the constituent waveforms in Fig. 5.24 could be clearly understood with the help of Fig. 5.25 and Table 5.2. The simple algebraic summation of the in-phase $s_{i1}\varphi_1(t)$ and quadrature $s_{i2}\varphi_2(t)$ waveforms generate the ultimate QPSK signal. This analysis is sufficient to explain all the waveform of Fig. 5.24.

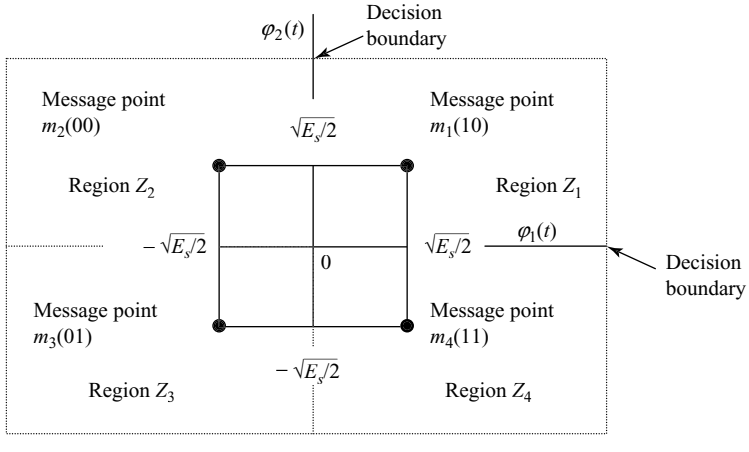


Fig. 5.26 *Signal Space diagram of coherent QPSK, M = 4*

Figure 5.26 is the constellation diagram for message points in QPSK showing the decision boundary regions. If a QPSK modulated signal undergoes filtering to reduce the spectral side lobes, the resulting

waveform will no longer have a constant envelope and in fact, the occasional 180^0 shifts in phase will cause the envelope to go to zero momentarily.

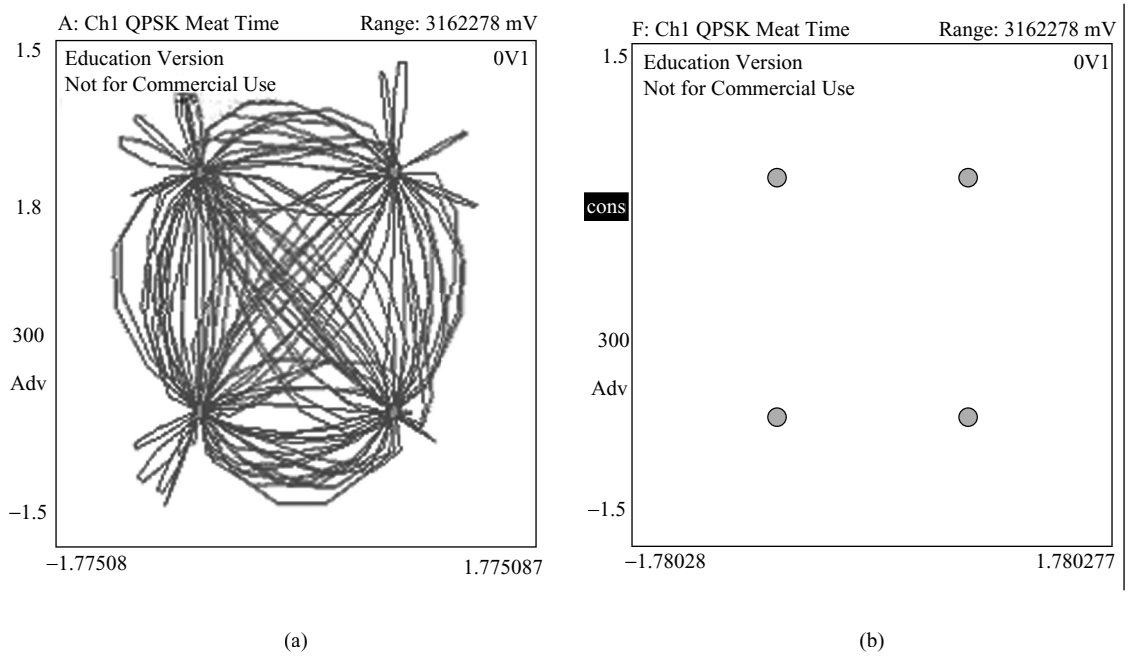


Fig. 5.27 *Measured signal space diagram of QPSK modulated carrier by Vector Signal Analyzer,* (*a*) *IQ diagram for complex recovered signal at the receiver in presence of noise,* (*b*) *contallation diagram for QPSK modulation with four message points*

Figure 5.27 is the signal space constellation diagram measured by Vector Signal Analyzer (VSA), which is elaborately explain in the later section. Figure 5.25 is a theoretical representation of the QPSK signal space diagram. In practical environment this space diagram is highly deviated from the theoretical one because of the channel noise, multipath fading, etc., and requires to be rectified through pulse shaping filter with proper roll of factor. Figure 5.27(a) is the representation of the IQ diagram under noisy environment with the symbol clock marked with small blue cloured circle. It shows the carrier's magnitude and phase synchronous with the symbol clock. This view gives some insight what actually the receiver sees the signal in order to demodulate the signal. Figure 5.27(b) is the constellation diagram for QPSK signal with four message points.

5.17.1 Error Probability of QPSK Signal

A QPSK signal has a two-dimensional signal constellation and four message points whose phase angles increase in a counterclockwise direction. Like binary PSK, the QPSK signal has minimum average energy. QPSK is basically equivalent to two binary PSK systems working in parallel using two carriers in phase quadrature. Using the signal–space diagram as shown in Fig. 5.26, it is observed that four message points in QPSK is circularly symmetric with respect to origin. So, the average probability of bit error for all message points is expressed in a different way in terms of separation distance between message points.

Considering a digital communication system in which two equally likely messages are represented by two vectors s_i and s_k. The decision boundary is represented by the bisector that is perpendicular to the line joining the points s_i and s_k. Accordingly, when the message m_i (vector s_i) is sent, and if the observation vector r lies on the side of the bisector where s_k lies an error is made. The probability of this event is

$$P(s_i, s_k) = P\ (r \text{ is closer to } s_k \text{ than } s_i, \text{ when } s_i \text{ is sent})$$

$$= \int_{d_{ik}/2}^{\infty} 1/(\pi N_0) \exp(-v^2 / N_0)\, dv \tag{5.78a}$$

where d_{ik} is the Euclidian distance between s_i and s_k. The complementary error function,

$$\text{erfc}(u) = 2/\sqrt{\pi} \int_{u}^{\infty} \exp(-z^2)\, dz \quad Let\ z = v/\sqrt{N_0} \text{ is substituted in Eq.(5.78a)}$$

Then, $$P(s_i, s_k) = 1/2\ \text{erfc}(d_{ik}/2\sqrt{N_0}) \tag{5.78b}$$

Considering all the message points m_i together, the probability of error $P_e(m_i)$ during the transmission of all message is represented as

$$P_e(m_i) \le 1/2 \sum_{\substack{k=1 \\ k \ne i}}^{M} \text{erfc}(d_{ik}/2\sqrt{N_0}) \quad i = 1, 2, 3. \ \&\ M \tag{5.79}$$

The probability of symbol error averaged over all the M message symbols is overbounded as

$$p_c = \sum_{i=1}^{M} p_i P_c(m_i)$$

$$\le 1/2 \sum_{i=1}^{M} \sum_{\substack{k=1 \\ k \ne i}}^{M} p_i\ \text{erfc}(d_{ik} / 2\sqrt{N_0}) \tag{5.80}$$

where p_i is the probability of transmitting symbol m_i. Now consider the symbol transmission is circularly symmetric about the origin. Then the conditional probability $P_e(m_i)$ is same for all i, in that situation Eq. (5.80) can be represented as

$$p_c \le 1/2 \sum_{\substack{k=1 \\ k \ne i}}^{M} \text{erfc}(d_{ik} / 2\sqrt{N_0}) \quad \text{for all } i, \tag{5.81}$$

Now, considering QPSK signal, let the message point m_1 corresponding to dibit 10 is transmitted. The closest two message points are m_2 and m_4 corresponding to 00 and 11 and they are equidistant from m_1.

So, $d_{12} = d_{14} = 2\sqrt{E_s}$

The message point m_3 is at more distant point with large E_s/N_0 value. From the observation vector in detecting message m_1, mistakenly message m_2 or m_4 can be considered. In doing so, a single bit error may occur. Again if m_3 is mistakenly chosen for m_1, there will be two bit errors. For large E_s/N_0, the chances of making two bits error is much less compared to 1 bit error. So, in calculating p_e, m_3 is excluded when message m_1 is sent.

The average probability of symbol error in each channel of a coherent QPSK system is then obtained from Eq. (5.81), by putting $d_{ik} = d_{12} = d_{14} = 2\sqrt{E_s}$ and $E_s = 2\ E_b$ as there are two bits per symbol.

$$p_c = \text{erfc}\left[\sqrt{E_b / N_0}\right] \tag{5.82}$$

and the Bit Error Rate (BER) for QPSK signals is

$$\text{BER}=1/2 \text{ erfc } (\sqrt{E_b /N_0}) \tag{5.83}$$

The in-phase and quadrature channels in QPSK system are statistically independent. In QPSK system there is two bits per symbol, so the transmitted signal energy per symbol is double the signal energy per bit. The average probability of bit error for coherent QPSK is same as that of binary PSK for the same bit rate and same E_b/N_0 but uses only half of the channel bandwidth. QPSK uses channel bandwidth better than binary PSK. The bandwidth of QPSK signal is half of the bandwidth of the BPSK signal. But the noise immunity in both schemes is same.

5.17.2 Generation and Detection of QPSK Signals

Figure 5.28 shows the generation and detection system of QPSK signals. The input binary data sequence is first converted into NRZ encoded data. Thus the symbol 1 and 0 is represented by $+\sqrt{E_s}$ and $-\sqrt{E_s}$. Using demultiplexer, two binary waves $g_1(t)$ and $g_2(t)$ are separated whose amplitudes are s_{i1} and s_{i2} respectively. Next $g_1(t)$ and $g_2(t)$ are modulated by two orthonormal basis functions

$$\varphi_1(t) = \sqrt{2/T_s} \cos (2\pi f_c t) \text{ and } \varphi_2(t) = \sqrt{2/T_s} \sin (2\pi f_c t)$$

As a result pair of binary PSK signals are produced, which are finally added to provide desired QPSK signal. This QPSK signal is transmitted over noisy AWGN channel and at the receiver observed signal $r(t)$ is received. Detector system consists of pair of correlators and arrangement for locally generated coherent reference $\varphi_1(t)$ and $\varphi_2(t)$. The correlator output r_1 and r_2 thus produced are compared at decision devices with threshold 0. If $r_1 > 0$, symbol 1 is detected otherwise symbol 0 is detected. Similarly, if $r_2 > 0$, a decision is made in favour of 1, otherwise decision is made for 0. The binary sequences produce in this manner in both in-phase and quadrature channels are finally combined using multiplexer to reproduce the original binary sequence at the transmitter input with minimum error probability.

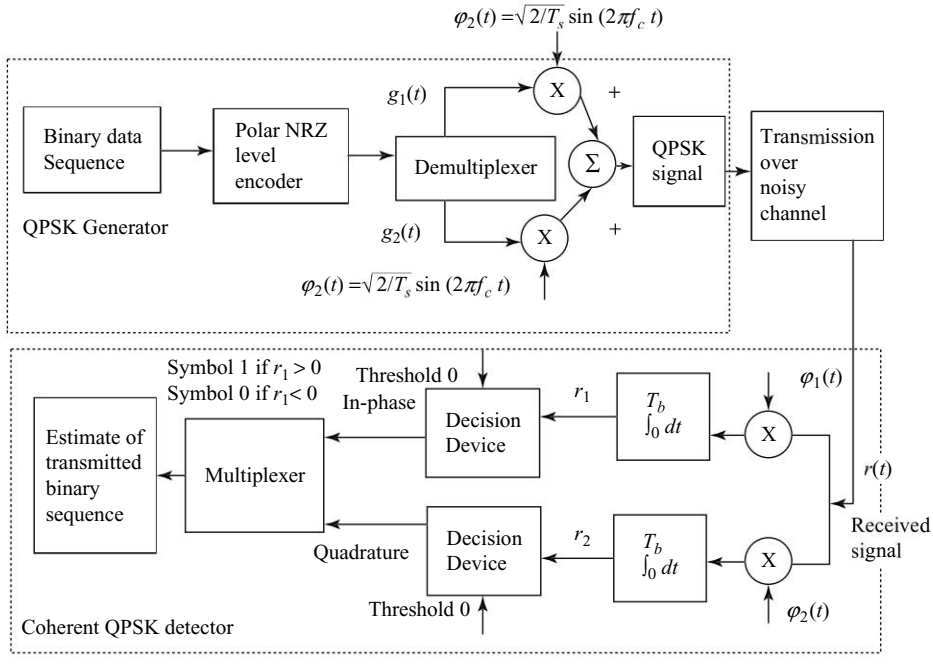


Fig. 5.28 *QPSK signal generation and detection system*

5.17.3 Testbed Implementation of QPSK Signal Using Two BPSK Modulators

The QPSK modulator consists of two BPSK modulators with two data inputs taken from two output of Word Generator (Binary Counter IC 74393), carriers phase shifted between the two streams (odd and even one) by 90°. We have taken one BPSK stream (considered as the odd stream), and added with the even stream to get QPSK output. Figure 5.29 shows the basic block diagram of generation of QPSK modulation using two BPSK. It consists of data generator, integrator, unipolar to bipolar converter (level shifter), multiplier and adder. Fig. 5.30 is the experimental testbed for QPSK modulator. Figure 5.31 is the experimentally observed bipolar data for both odd and even stream.

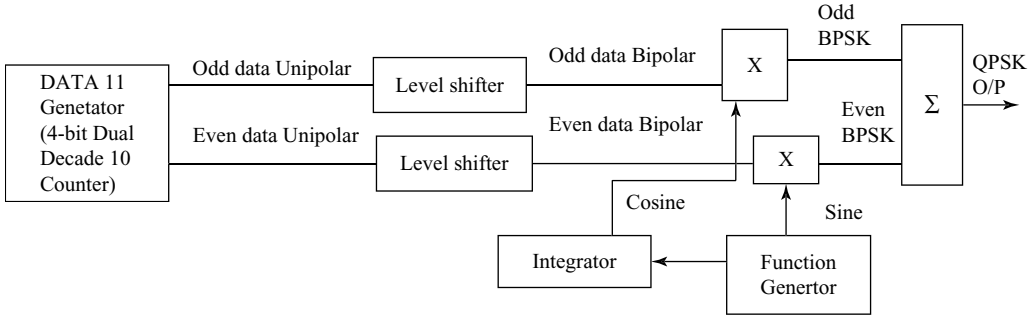
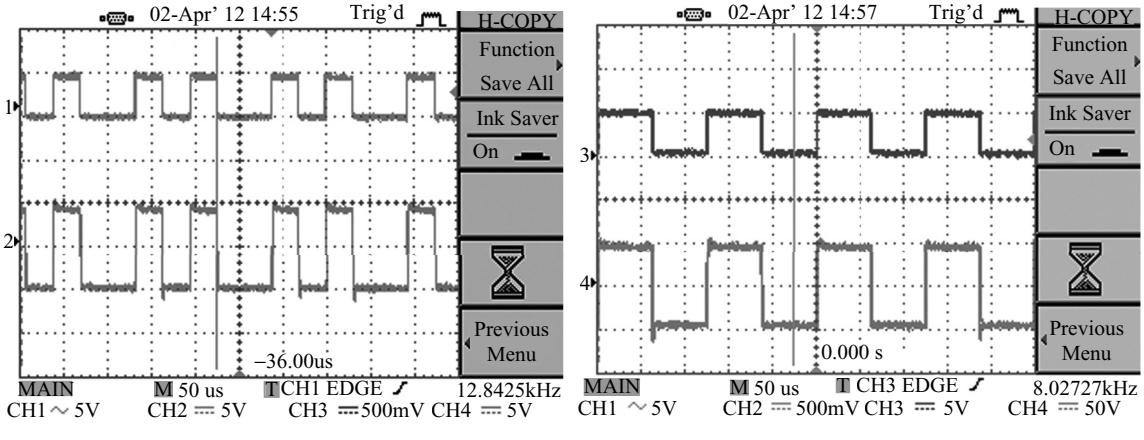


Fig. 5.29 *Block diagram for QPSK modulator tesTbed*



(a) (b)

Fig. 5.30 *Unipolar to bipolar conversion of data for* (*a*) *Odd data stream,* (*b*) *Even data stream*

Figure 5.32 is the experimental observed results through 4- channel digital storage oscilloscope for odd, even BPSK and QPSK modulated signal. It shows the correct waveforms in accordance to the theoretical explanation given in this chapter.

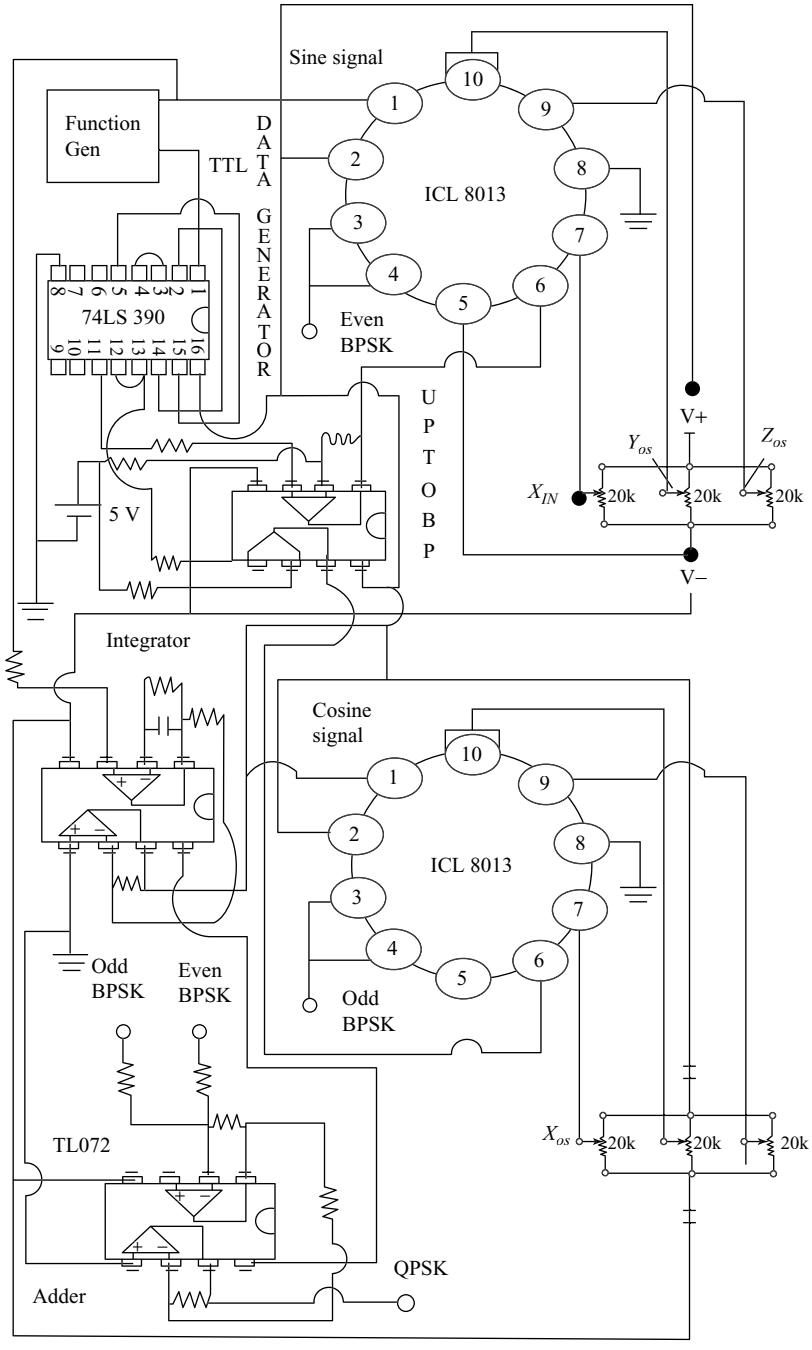


Fig. 5.31 *Testbed for QPSK modulator*

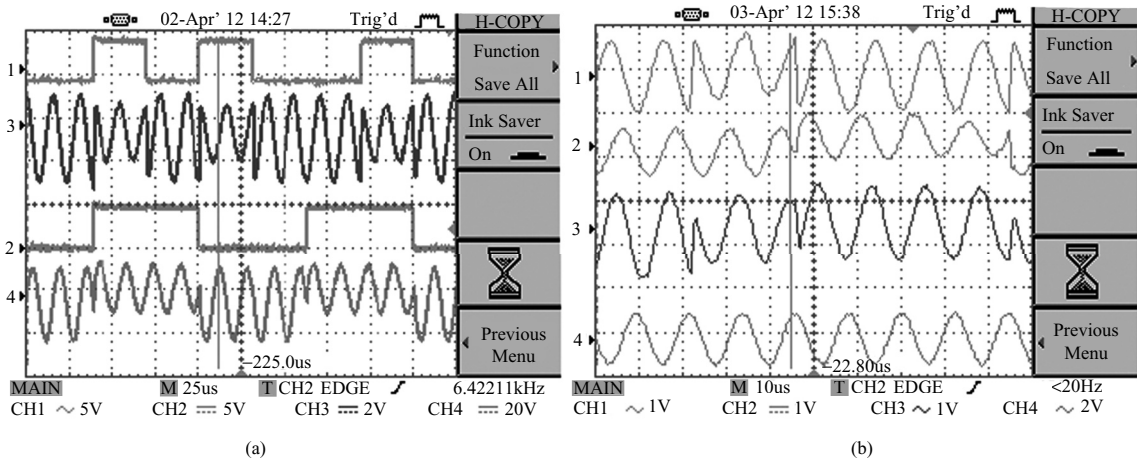


Fig. 5.32 *QPSK testbed output (a) Ch 1 odd data, Ch 2 even data, Ch 3 odd BPSK, Ch 4 even BPSK, (b) Ch 1 odd BPSK, Ch 2 even BPSK, Ch 3 QPSK, Ch 4 Carrier*

5.17.4 Power Spectra of QPSK Signals

In the BPSK waveform, there are rectangular pulses of amplitude $\pm\sqrt{2E_b/T_b}$ over the duration $0 \le t \le T_b$ depending on whether we have symbol 1 or 0. In QPSK, depending on the dibit sent during the signaling interval T_s, the amplitude of in-phase component equals to $+g(t)$ or $-g(t)$, similarly for quadrature component. So, the symbol shaping function $g(t)$ of the QPSK signal can be defined as

$$g(t) = \begin{matrix} \sqrt{E_s/T_s} & 0 \le t \le T_s \\ 0 & \text{otherwise} \end{matrix} \tag{5.84}$$

Assuming the binary wave at the modulator input is random and symbol 1 and 0 are equally likely and transmission at two adjacent time slots is statistically independent, both the in-phase and quadrature components have common power spectral density each given as $E_s\,\text{sinc}^2(T_s f)$. So, the power spectral density of QPSK signal is the sum of the two components as they are statistically independent components.

$$S_{\text{QPSK}} = 2E_s\,\text{sinc}^2\,(T_s f) = 4\,E_b\,\text{sinc}^2(2T_b f) \tag{5.85}$$

This also comes from Eq. (5.71) for MPSK by putting $M = 4$.

5.17.5 Offset Quadrature Phase Shift Queuing (OQPSK)

If the two bit streams I and Q are offset by a 1/2 bit interval, then the amplitude fluctuations are minimized since the phase never changes by 180° as it is occurred in QPSK when dibit 01 changes to 10. This modulation scheme, Offset Quadrature Phase shift Keying (OQPSK) is obtained from QPSK by delaying the odd bit stream by half a bit interval with respect to the even bit stream as shown in Fig. 5.33.

Thus the range of phase transitions is between 0° and 90° (the possibility of a phase shift of 180^0 is eliminated as occurred in QPSK at position 2T) and occurs twice as often, but with half the intensity of the QPSK as shown in Fig. 5.32(b). While amplitude fluctuations still occur in the transmitter and receiver they have smaller magnitude. The bit error rate for QPSK and OQPSK are the same as for BPSK. When an OQPSK signal undergoes band limiting, the resulting intersymbol interference causes the envelop to droop slightly to the region of ± 90° phase transition, but since the phase transitions of 180° have been avoided in OQPSK, the envelop will never go to zero as it does in QPSK.

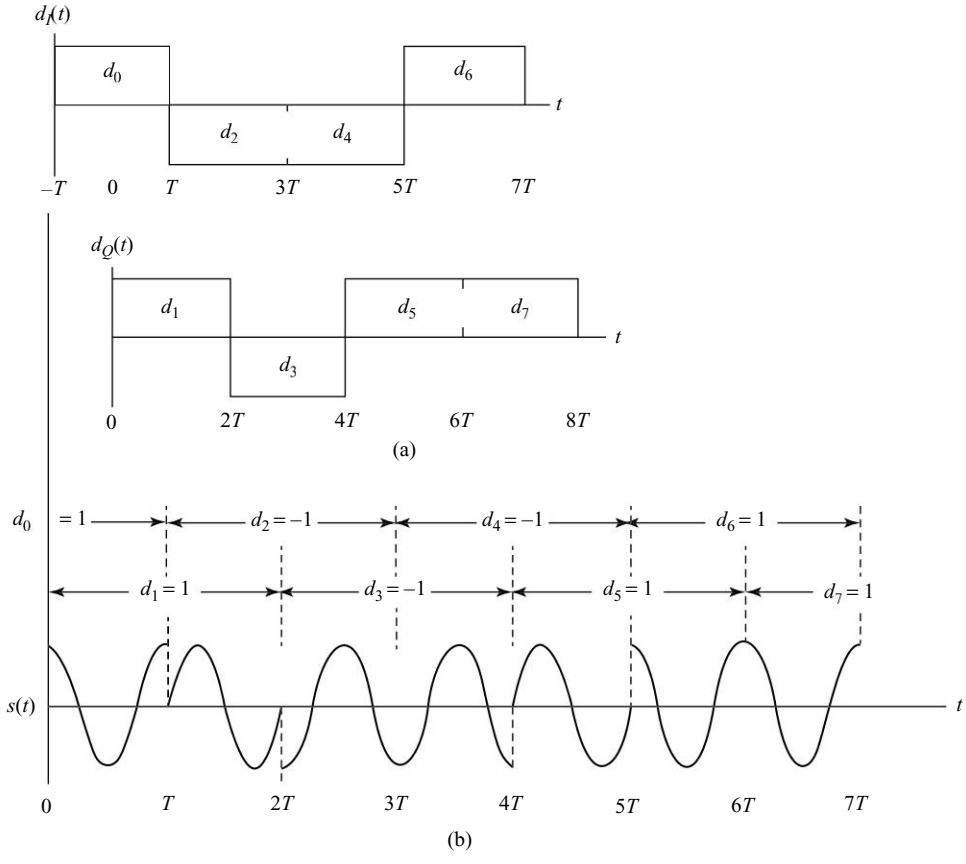


Fig. 5.33 *Generation of OQPSK signal (a) odd and even bit stream, (b) OQPSK signal*

5.18 M-ARY QUADRATURE AMPLITUDE MODULATION

Unlike the M-ary PSK modulation where amplitude remains constant both in in-phase and quadrature components, in M-ary Quadrature Modulation (MQAM), the carrier exhibits both the amplitude and phase modulation. So, this is called *hybrid modulation technique*. The M-ary Pulse Amplitude Modulation (PAM) is a one-dimensional scheme, whereas MQAM is a two-dimensional modulation scheme where two orthogonal basis functions are used for the formation of the MQAM signals. The basis function is given as

$$\varphi_1(t) = \sqrt{2/T_s}\cos(2\pi f_c t) \text{ and } \varphi_2(t) = \sqrt{2/T_s}\sin(2\pi f_c t) \text{ over } 0 \le t \le T_s$$

Within the $(\varphi_1(t), \varphi_2(t))$ plane, the message point s_i is represented by $(m_i d_{min}/2, n_i d_{min}/2)$, where d_{min} is the minimum distance between two message points in the constellation, m_i, n_i are integers, $i = 1,2,3, \ldots M$. If E_0 is the minimum energy associated with the signal having lowest amplitude, then $d_{min}/2 = \sqrt{E_0}$
Thus the MQAM signal for any symbol k, can be expressed as

$$\begin{aligned} s_k(t) &= \sqrt{2E_0/T_s}\, m_k \cos(2\pi f_c t) - \sqrt{2E_0/T_s}\, n_k \sin(2\pi f_c t) \\ &= p(t)(m_k \cos(2\pi f_c t) - n_k \sin(2\pi f_c t)) \\ &= r_k\, p(t) \cos(2\pi f_c t - \theta_k) \end{aligned} \quad (5.86)$$

Over $0 \le t \le T_s, k = 0, \pm 1, \pm 2, \ldots$

where $p(t) = \sqrt{2E_0 / T_s}$ = Properly shaped base band pulse, $r_k = \sqrt{m_k^2 + n_k^2}$,

$\theta_k = \tan^{-1}(n_k/m_k)$.

The signal $s_k(t)$ consists of two phase–quadrature carriers with each one being modulated by a set of discrete amplitudes, hence the name Quadrature Amplitude Modulation (QAM). Depending on the number of possible symbols M, two distinct constellations for QAM are obtained. One is the Square Constellation for which the number of bits per symbol is even and two, the Cross Constellation in which the number of bits per symbol is odd. In square constellation, an M-ary QAM quare constellation can always be represented by the Cartesian product of one-dimensional L-ary Pulse Amplitude Modulation (PAM) constellation with itself, where $L = \sqrt{M}$. To generate M-ary QAM signal with odd number of bit sequences per symbol, it requires cross constellation. It is not possible to express a QAM cross constellation as the product of a PAM constellation with itself.

With $M = 4$ (4-ary or quaternary case), we have four basic symbols or pulses. A sequence of binary digits can be transmitted by 4-ary symbols as shown in Fig. 5.34. This is because two binary digits can form four possible combinations of binary sequences 11, 10,01,00. This type of signaling is also called multi-amplitude signaling. To transmit n binary digits, we need only $(n/2)$ 4-ary pulses. A group of 3 bits can be transmitted by one 8-ary symbol. In general, the information I_M transmitted by an M-ary symbol is

$$I_M = \log_2 M \text{ binary digits or bits} \tag{5.87}$$

This implies the rate of information transmission increases as M increases at the cost of increase transmitted power.

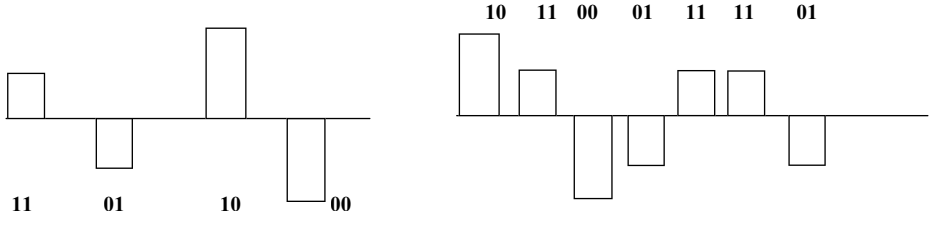


Fig. 5. 34 *Example of 4-ary multi-amplitude signal*

5.18.1 16-ary QAM or 16 QAM

From Eq. (5.86), a signal $p_k(t)$ can be generated using QAM by letting $m_1(t) = m_k\, p(t)$ and $m_2(t) = n_k\, p(t)$. The transmitted pulse $p_k(t)$ can take on 16 distinct forms and is therefore 16-ary pulse as shown in Fig. 5.35. This method of generation pulses is called 16-QAM signals for which $M = 16$, each encoded message points needs 4 bits, i.e each pulse can transmit the information of $\log_2 16 = 4$ binary digits. There are 16 possible sequences of four binary digits and there are 16 combination of (m_k, n_k). Thus every possible 4-bit sequence is transmitted by a particular (m_k, n_k) or (r_k, θ_k). Therefore each signal pulse $r_k\, p(t)\cos(2\pi f_c t - \theta_k)$ transmits 4 bits. Hence, bit rate of transmission increases without increasing bandwidth.

In the encoding process two leftmost bits represent the quadrant in (φ_1, φ_2) plane and the other two bits are used to represent one of the four possible symbols lying within each quadrant of the (φ_1, φ_2) plane, starting from first quadrant to the fourth in counter clockwise direction as (11, 10, 00, 01).

In the square matrix representation, $\{m_k, n_k\}$ can be given as follows.

$$\{m_k = n_k\}\begin{bmatrix} (-3,3) & (-1,3) & (1,3) & (3,3) \\ (-3,1) & (-1,1) & (1,1) & (3,1) \\ (-3,-1) & (-1,-1) & (1,-1) & (3,-1) \\ (-3,-3) & (-1,-3) & (1,-3) & (3,-3) \end{bmatrix} \tag{5.88}$$

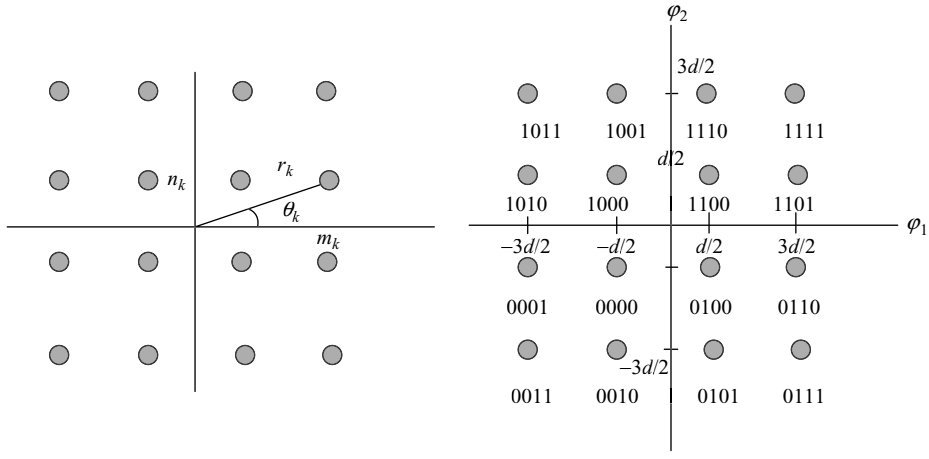


Fig. 5.35 *Signal–space representation of 16-Ary QAM following Gray-Encoded quadbits*

Thus the square constellation of Fig. 5.35 is the Cartesian product of the 4-PAM constellation as shown in Fig. 5.36 with itself. The representation of the four possible dibits based on Gray encoding for 4-PAM are also shown in the figure.

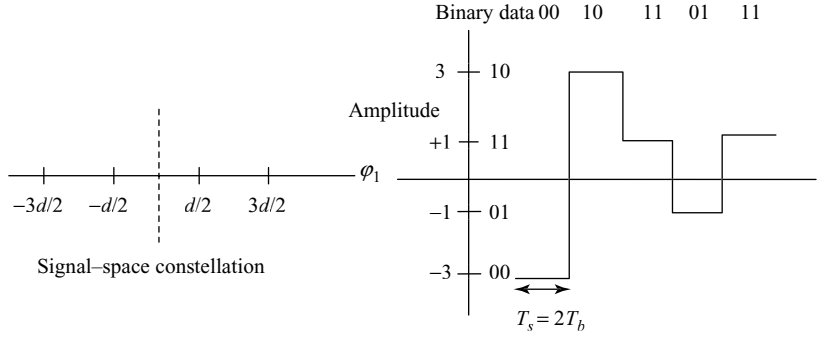


Fig. 5.36 *4-PAM constellation and waveform with the representation of 4-possible dibits*

As the M-ary square constellation can be represented as the Cartesian product of one-dimensional L-ary (L amplitude levels) PAM constellation with itself, where $L=\sqrt{M}$. In that case the general representation of QAM square constellation is a square matrix of the form given below.

$$\{m_k = n_k\} = \begin{bmatrix} (-L+1,\ L-1) & (-L+3,\ L-1)\ldots & (L-1,\ L-1) \\ (-L+1,\ L-3) & (-L+3, L-3)\ldots & (L-1,\ L-3) \\ \vdots & \vdots & \vdots \\ (-L+1,\ -L+1) & (-L+3, -L+1)\ldots & (L-1, -L+1) \end{bmatrix} \tag{5.89}$$

5.18.2 Error Probability of M-ary QAM

Let the probability of symbol error for L-ary PAM with $L = \sqrt{M}$ is $p_{e,}$

$$p_e = (1 - 1/\sqrt{M})\ \text{erfc}\ (\sqrt{E_0 / N_0}) \tag{5.90}$$

The probability of correct detection of M-ary QAM $p_c = (1 - p_{e,}^{'})\ (1 - p_e^{'})$

The probability symbol error for M-ary QAM $p_e = 1 - p_c \cong 2\, p_e^{'}$

$$p_e = (1-1/\sqrt{M})\ \text{erfc}\ (\sqrt{E_0/N_0}) \tag{5.91}$$

$$P_e = (1-1/\sqrt{M})\ \text{erfc}\ (d_{\min}/2\sqrt{N_0})$$

E_0 is the minimum energy associated with lowest amplitude of the transmitted signal.

For M-ary QAM, the transmitted signal energy is variable depending on the symbol amplitude. Therefore, the symbol error probability is expressed in terms of average value of the transmitted energy rather than E_0. Considering the amplitude levels of inphase and quadrature components are equally likely, the average energy is calculated as,

$$E_{\text{av}} = 2\ [2E_0/L \sum_{i=1}^{L/2} (2i-1)^2] \tag{5.92}$$

The factor 2, is for two components inphase and quadrature. After performing the summation operation, $E_{av} = 2(M–1)\ E_0/3$, putting this value we get the expression for p_e finally as

$$p_e = 2\ (1-1/\sqrt{M})\ \text{erfc}\ \left(\sqrt{3E_{av}\ /\ (2(M-1)N_0}\right) \tag{5.93}$$

5.19 COHERENT FREQUENCY SHIFT KEYING (FSK)

We have already discussed the linear modulation of M-ary PSK and M-ary QAM. Coherent FSK is the non-linear method of pass band transmission. At first we shall discuss the binary FSK system then the Minimum Phase Shift Keying will be discussed.

5.19.1 Binary FSK

In a binary FSK system, symbols 1 and 0 are distinguished from each other by transmitting one of two sinusoidal waves that differ in frequency by a fixed amount. A typical pair of sinusoidal waves is described by

$$s_i(t) = \begin{cases} \sqrt{2E_b/T_b}\ \cos(2\pi f_i t) \\ 0,\ \text{elsewhere} \end{cases} \tag{5.94}$$

$$i = 1,2\ ,\ 0 \le t \le T_b$$

E_b is the transmitted signal energy /bit, and $f_i = (n + i)\ /T_b$, for some fixed integer n.

Thus symbol 1 is represented by $s_1(t)$ and symbol 0 by $s_2(t)$. The phase continuity of the signals is always maintained along with inter-bit switching times. This is known as *continuous phase frequency shift keying* (CFSK). The orthonormal basis function is therefore given as

$$\varphi_i(t) = \begin{cases} 2/T_b \cos(2\pi f_i t) & i = 1,2\ ,0 \le t \le T_b \\ 0,\ \text{otherwise} \end{cases} \tag{5.95}$$

The coefficient s_{ij} for $i = 1,2$ and $j = 1,2$ is defined as

$$s_{ij} = \int_0^{T_b} s_i(t)\varphi_j(t)\,dt = \begin{cases} \sqrt{E_b} & i = j \\ 0 & i \ne j \end{cases}$$

Unlike binary PSK, a binary FSK system is characterized by two-dimensional vector space ($N = 2$) with two message points ($M = 2$) as shown in Fig. 5.37. The two message points are defined as

$$s_1 \begin{bmatrix} \sqrt{E_b} \\ 0 \end{bmatrix} \qquad\qquad s_2 \begin{bmatrix} 0 \\ \sqrt{E_b} \end{bmatrix}$$

The Euclidean distance between them is equal to $\sqrt{2E_b}$

5.19.2 Error Probability of BFSK Signals

The observation vector r has two elements corresponding to two message points on the signal-space diagram. Let r_1 and r_2 correspond these two elements as defined below.

$$r_1 = \int_0^{T_b} r(t)\varphi_1(t)\,dt \tag{5.96}$$

$$r_2 = \int_0^{T_b} r(t)\varphi_2(t)\,dt$$

where r(t) is the received signal. When symbol 1 is transmitted, r(t) equals $s_1(t) + w(t)$ and when symbol 0 is transmitted, $r(t)$ equals $s_2(t) + w(t)$.

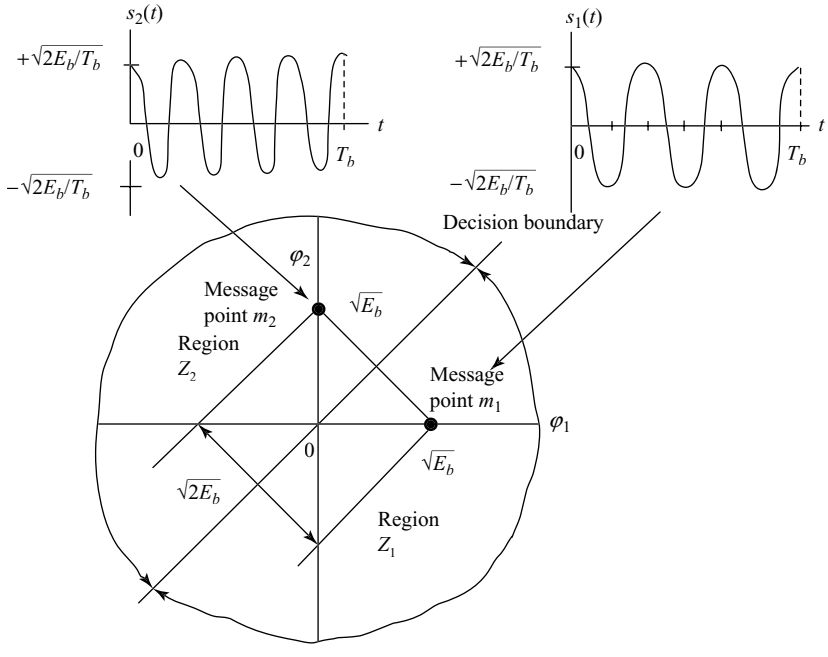


Fig. 5.37 *Signal-space diagram for BFSK signals*

Where $w(t)$ is the Gaussian noise with spectral density $N_0/2$. The receiver decides in favour of symbol 1 if the received signal point represented by the observation vector r falls inside region Z_1, this occurs $r_1 > r_2$ and if $r_1 < r_2$, the received signal point falls inside region Z_2, receiver decides in favour of symbol 0. In the decision making process, error may occur, that depends on the Euclidean distance as given in Eqs. (5.78a–5.78b). For BFSK, the Euclidean distance is $\sqrt{2E_b}$. Considering Eq. (5.74a) and putting the value for d_{ik} as $\sqrt{2E_b}$,

$$P(s_i, s_k) = P\ (r \text{ is closer to } s_k \text{ than } s_i, \text{ when } s_i \text{ is sent})$$

$$= 1/2\ \text{erfc}\ (d_{ik}/2\sqrt{N_0}) \ = \ 1/2\ \text{erfc}\ (\sqrt{E_b/2N_0}\ )$$

This is to be mentioned that for BPSK the Euclidean distance is $2\sqrt{E_b}$ and the error probability is $1/2\ \text{erfc}\left(\sqrt{(E_b/N_0}\right)$ which is in perfect accordance with the two cases.

5.19.3 Generation and Detection of Coherent Binary FSK Signals

To generate BFSK signals, the binary sequence is first applied for on-off level encoder. At the output of the encoder an equivalent voltage of amplitude $\sqrt{E_b}$ volts is produced corresponding to symbol 1 and 0 volt corresponding to symbol 0. Two channels are used in which one channel is associated with frequency f_1 and the other channel with frequency f_2. One inverter is used at the lower channel to invert the voltage produced at upper channel. In that way, it ensures that f_1 frequency modulation happened when symbol 1 is transmitted, and f_2 frequency modulation happened when symbol 0 is transmitted. Summing up the two modulated signals, BFSK signals will be obtained. This is shown in Fig. 5.38. The two frequencies f_1 and f_2 are chosen such to equal different integer multiples of the bit rate $1/T_b$ i.e., $f_i = (n + i)/T_b$.

At the detector circuit two correlators with common input signal $r(t)$ is used along with locally generated referenced signals $\varphi_1(t)$ and $\varphi_2(t)$. The correlators output are then subtracted one from the other to get a difference to be compared with threshold (0). If the difference is > 0, symbol 1 is decided, if <0, symbol 0 is decided. This is also shown in Fig. 5.38.

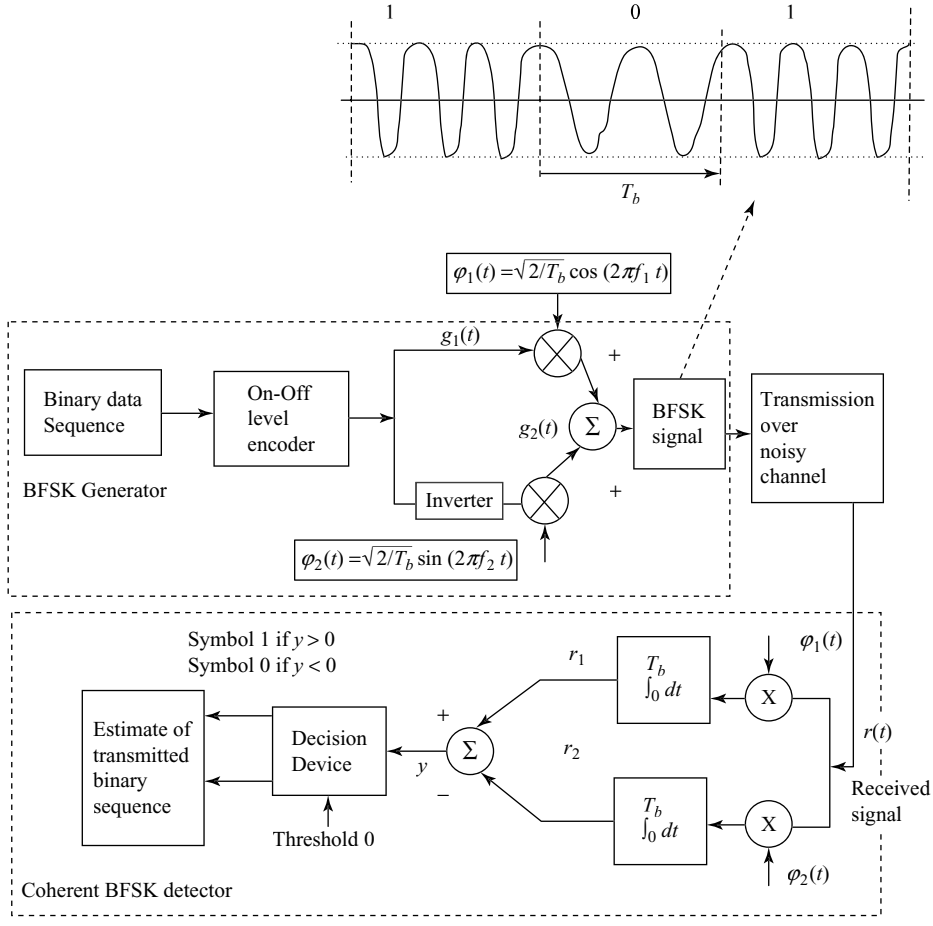


Fig. 5.38 *Generation and detection systems of BFSK*

5.19.4 Power Spectra of BFSK Signal

In Sunde's FSK, the transmitted frequencies f_1 and f_2 differ by an amount equal to the bit rate $1/T_b$ and their arithmetic mean is equal to the carrier frequency f_c. Phase continuity is always maintained. The binary FSK signal can be expressed as,

$$s(t) = \sqrt{2E_b/T_b}\cos(2\pi f_c t \pm \pi t/T_b), \qquad 0 \le t \le T_b$$
$$= \sqrt{2E_b/T_b}\{\cos(2\pi f_c t)\cos(\pm \pi t/T_b) - \sin(2\pi f_c t)\sin(\pm \pi t/T_b)\}$$
$$= \sqrt{2E_b/T_b}\{\cos(2\pi f_c t)\cos(\pi t/T_b) - \sin(2\pi f_c t)\sin(\pi t/T_b)\} \quad (5.99)$$

In Eq. (5.99), + sign is for transmitting symbol 0 and – sign is for transmitting symbol 1. Symbols 1 and 0 are randomly generated wave at the modulator input that is equally likely and symbols transmission in two adjacent time slots are statistically independent. Under such assumption, the BFSK signal can be written as

$$s(t) = \sum_{k=-\infty}^{\infty} g(t - kT_b)\{\cos(2\pi f_c t + I_k \pi t/T_b)\} \quad (5.100)$$

where $I_k = \pm 1$ with equal probability, and

the pulse shaping function $g(t) = \begin{cases} \sqrt{2E_b/T_b}, & 0 \le t \le T_b \\ 0 & \text{otherwise} \end{cases}$

Let the in-phase and quadrature components of $s(t)$ are defined as

$$s(t) = \sum_{k=-\infty}^{\infty} g(t - kT_b)\,\text{Re}\{\exp j(2\pi f_c t + I_k \pi t/T_b)\} \quad (5.101)$$

$$s(t) = \sum_{k=-\infty}^{\infty} g(t - kT_b)\left\{e^{jI_k \pi t/T_b}\right\} e^{j2\pi f_c t}$$
$$= \left\{g_I(t) + jg_Q(t)\right\} e^{j2\pi f_c t} \quad (5.102)$$

where $g_I(t)$ = in-phase component

$$= \sum_{k=-\infty}^{\infty} g(t - kT_b)\cos(\pi t/T_b) = \sqrt{2E_b/T_b}\cos \pi t/T_b \quad (5.103)$$

Which is completely independent of input binary wave.
$g_Q(t)$ = quadrature component

$$= \sum_{k=-\infty}^{\infty} I_k g(t - kT_b)\sin(\pi t/T_b) = \pm\sqrt{2E_b/T_b}\sin \pi t/T_b \quad (5.104)$$

The quadrature component is directly related with input wave. Within the symbol interval $0 \le t \le T_b$, it is equal to $-g(t)$ for symbol transmission 1 and $+g(t)$ for symbol 0.
The autocorrelation function of $g_I(t)$

$$R_{gIgI}(\tau) = (1/T_b)\int_0^{T_b} (2E_b/T_b)\cos(\pi(t + \tau/T_b)\cos(\pi t/T_b) = (E_b/T_b)\cos(\pi t/T_b) \quad (5.105)$$

The power spectrum of this component is

$$S_{gI}(f) = \text{Fourier transform of } R_{gIgI}(\tau) = \int_0^{T_b} (E_b/T_b)\cos(\pi t + T_b)\, e^{-j2\pi ft}\, dt$$
$$= (E_b/2T_b)\,[\delta(f - 1/2T_b) + \delta(f + 1/2T_b)\,] \quad (5.106)$$

The Fourier transform of the $G_Q(t) = G_Q(f) = \int_0^{T_b} \sqrt{2E_b/T_b}\ \sin(\pi t/T_b) e^{-j2\pi ft}\, dt = 1$

Integrating by parts with the limit,

$$G_Q(f) = 2\sqrt{2E_b/T_b}\ /\ \pi\ [\cos(\pi f T_b)/1 - 4T_b^2 f^2] e^{(-j2\pi ft)} \tag{5.107}$$

The power spectra for quadrature components $S_{gQ}(f) = (1/T_b)\ G_Q(f) G_Q(-f)$

$$= 4\,x\,(2E_b)/\pi^2 [\cos((\pi f T_b)/1 - 4T_b^2 f^2]^2 \tag{5.108}$$

Therefore, the power spectra for BFSK is the sum of the $S_{gI}(f)$ and $S_{gQ}(f)$

$$S_{BFSK}(f) = (E_b/2T_b)\ [\delta(f - 1/2T_b)\ + \delta(f + 1/2T_b)] + (8E_b/\pi^2)\ [\cos((\pi f T_b)/1 - 4T_b^2 f^2]^2 \tag{5.109}$$

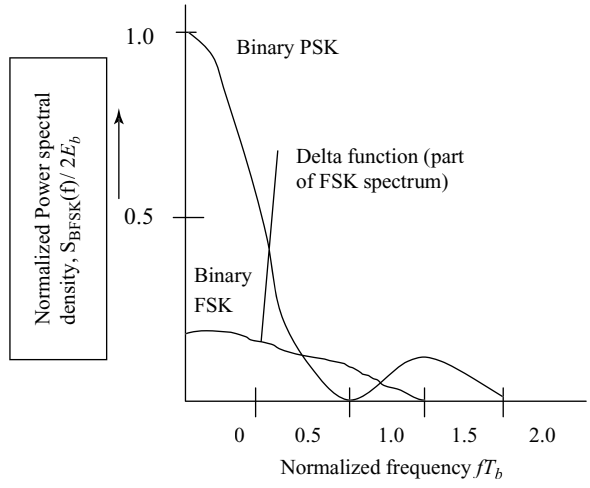


Fig. 5.39 *Normalized power spectra of BFSK and BPSK*

Thus the power spectra of binary FSK has two frequency components $f_1 = f + 1/2T_b$ and $f_2 = f - 1/2T_b$, with their average power adding up to one-half of the total power of binary FSK. The base band power spectral density of BFSK falls of as inverse of the fourth power of frequency f. The normalized power spectral of BFSK along with BPSK is shown in Fig. 5.39, it is seen that as pulse becomes smoother, faster roll off occurs.

If BFSK is not continuous phase, that is f_1 and f_2 are not multiple of $1/T_b$, then spectral roll off (rate of decay) is slower, in fact it decays as $1/f^2$ than that of $1/f^4$ as in continuous phase BFSK.

5.20 MINIMUM SHIFT KEYING: MSK

In the coherent BFSK signal, the phase information is used only to provide for synchronization of the receiver to the transmitter. This phase information is not fully exploited for improving the noise performance of the receiver significantly at the expense of receiver complexity. MSK is a special type of continuous phase FSK (CPFSK) where the frequency changes occur at the zero crossings of the carrier. The modulation index for MSK is 0.5. The modulation index of 0.5 corresponds to the minimum frequency spacing that makes two FSK signals to be coherently orthogonal and the name minimum shift keying implies the minimum frequency separation, i.e., bandwidth that allows orthogonal detection. It is found that binary data has sharp transition

between "1" and "0" states and vice versa. This creates the signal having side bands extending out of the carrier signal and causes interference to adjacent channels. By using filter, this problem can be partly overcome.

MSK has one of two possible frequencies over any symbol interval. In traditional FSK, we use signals of two different frequencies of f_1 and f_2 to transmit message $s = 0$ and message $s = 1$over a time T_b.

$$s_1(t) = \sqrt{2E_b / T_b}\ \cos\ (2\pi f_1 t)\ , \quad 0 \le t \le T_b$$
$$s_2(t) = \sqrt{2E_b / T_b}\ \cos\ (2\pi f_2 t), \quad 0 \le t \le T_b$$

We assume $f_1 > f_2 > 0$. Choice of frequencies are such that in each time interval T_b, there is an integer number of periods, $f_1 = k_1/T_b$ and $f_2 = k_2/T_b$ with k_1 and k_2 being integers, the signal is definitely continuous phase FSK. Figure 5.40 is an example of a signal that is discontinuous, a signal with discontinuous phase and a signal with continuous phase. As phase-continuous signals in general have better spectral properties than signals that are phase discontinuous, we prefer to transmit signals having this property of phase continuity.

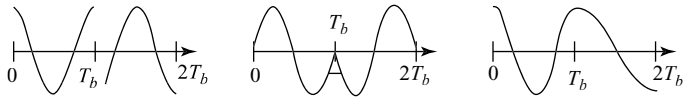


Fig. 5.40 *Signals with different degrees of discontinuity*

If either f_1 or f_2 are chosen such that there is non-integer number of periods, the traditional FSK modulator will output a signal with significant discontinuities in the phase. In order to maintain the phase continuity, we can let the transmitter to have memory. We choose the signals for a general CPFSK as

$$s(t) = \begin{cases} \sqrt{2E_b / T_b}\ \cos\ [\ 2\pi f_1 t\ + \theta(0)] \ \text{ for symbol } \ 1 = s_1(t) \ \ 0 \le t \le T_b & (5.110) \\ \sqrt{2E_b / T_b}\ \cos\ [\ 2\pi f_2 t + \theta(0)] \ \text{ for symbol } \ 0 = s_2(t) \ \ 0 \le t \le T_b & (5.111) \end{cases}$$

We keep the phase continuity by letting $\theta(0)$ be equal to the argument of the cosine pulse for the previous bit interval. For the signals over an arbitrary bit interval $kT_b \le t \le (t+1)T_b$, the general phase memory term is $\theta\ (kT_b)$. In Fig. 5.41, the phase variation is depicted by bold black line over time in a phase trellis for binary sequence 11000. We assume modulation index $h = 0.5$ and $\theta(0) = 0$ or $\theta(0) = \pi$. We see that for every multiple of the bit time the phase can only take on one of two values, the values being 0 and π for $t = 2kT_b$ and $\pm\ \pi/2$ for $t = (2k+1)T_b$. Thus CPFSK with deviation ratio $h = 0.5$ is called MSK. The frequency difference $f_d = f_1 - f_2 = 1/2T_b$ that results from choosing $h = 0.5$ is the smallest possible difference if the signals of the two frequencies are to be orthogonal over one bit interval.

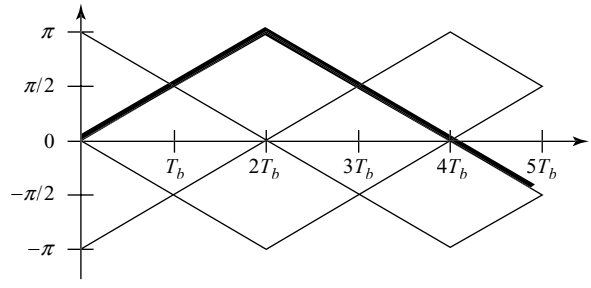


Fig. 5.41 *Phase trellis with h = 0.5, sequence 11000*

An example of an MSK signal with $k_1 = 1$ ($f_1 = 1/T_b$) and $k_2 = 1/2$ ($f_2 = 1/2T_b$) is given in Fig. 5.42.

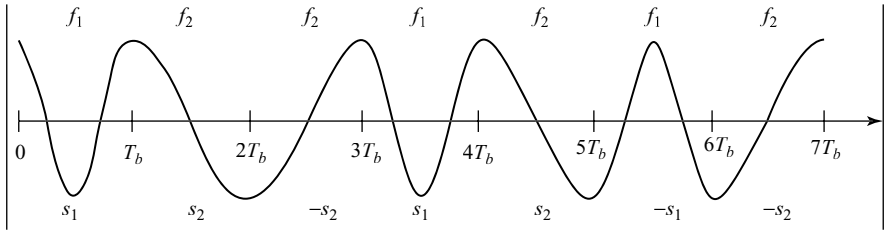


Fig. 5.42 *Example of MSK signal (CPFSK)*

MSK is unique due to the relationship between the frequencies of logic 0 and 1. The difference between the frequencies is always 1/2 the data rate. This is the minimum frequency spacing that allows 2 FSK signals to be coherently orthogonal that is why the name minimum shift keying.

The base band modulation starts with a bit stream of 0's and 1's and a bit-clock. The base band signal is generated by first transforming the 0/1 encoded bits into –1/1 using an NRZ filter. This signal is then frequency modulated to produce the complete MSK signal (Fig. 5.43).

The amount of overlap that occurs between bits will contribute to the inter-symbol interference (ISI).

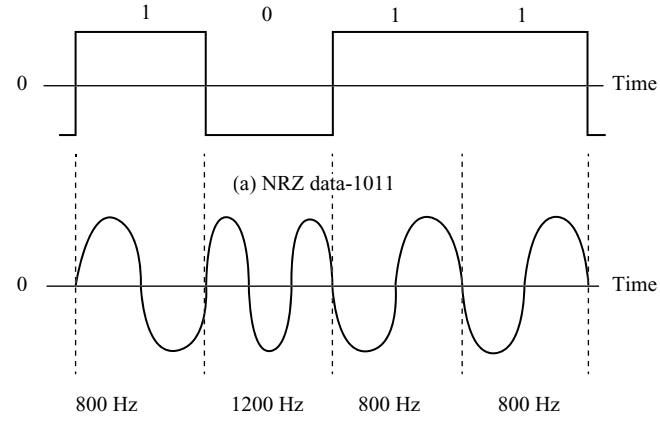


Fig. 5.43 *Example of MSK signal with frequency spacing 400 Hz (800 bits/sec base band MSK)*

In general the angle modulated wave for MSK is represented as

$$s(t) = \sqrt{2E_b / T_b}\ \cos\ [2\pi f_c t + \theta(t)] \tag{5.112}$$

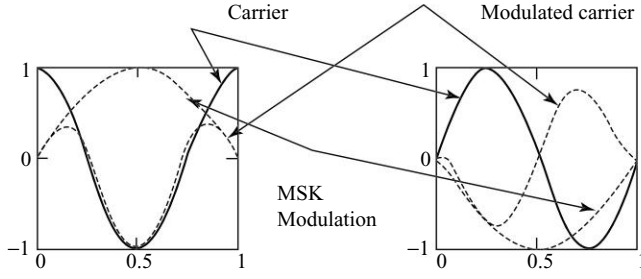
The carrier frequency $f_c = ½\ (f_1 + f_2)$ and $\theta\,(t)$ is the phase of $s\,(t)$, a continuous function of time. The deviation ratio h is defined with respect to T_b, the bit rate as

$$\begin{aligned} \theta(t) &= \theta(0) \pm \pi h t / T_b \qquad 0 \le t \le T_b \\ h &= T_b\,(f_1 - f_2) = T_b f_d \\ \theta(t) &= \theta(0) \pm \pi f_d t \end{aligned} \tag{5.113}$$

The + sign corresponds to the frequency of the carrier being shifted to higher frequency $f_2 = f_c + 1/4T_b$ while for – sign the frequency of carrier f_c shifs to lower frequency $f_1 = f_c - 1/4T_b$. So the difference $f_2 - f_1 = 1/2T_b$.

With $h = 0.5$, the frequency of separation is equal to half the bit rate. We previously showed that OQPSK is obtained from QPSK by delaying the Q data stream by 1 bit or T seconds with respect to I data stream. This delay has no effect on the error or bandwidth. MSK is derived from OQPSK by replacing the rectangular pulse in amplitude with a half-cycle sinusoidal pulse.

The MSK changes the pulse shape to a half cycle sinusoid. Figure 5.44 shows a MSK pulse signal and then multiplication by a carrier. The modulated carrier obtained as a result of multiplication of the pulse shape and the carrier.



(a) MSK pulse and carrier for a 1 bit (b) MSK pulse and carrier for a 0 bit

Fig. 5.44 *MSK pulse shaping is a half wave sinusoid*

Using the well-trigonometric identity in Eq. (5.112), $s(t)$ can be expressed as

$$s(t) = \sqrt{2E_b / T_b}\cos(2\pi f_c t)\cos(\theta(t)) - \sqrt{2E_b / T_b}\sin(2pf_c t)\sin(\theta(t)) \quad (5.114)$$
$$= s_I(t)\cos(2\pi f_c t) - s_Q(t)\sin(2\pi f_c t)$$

where $s_I(t)$ and $s_Q(t)$ are the in-phase and quadrature components of the MSK signal. With $h = 0.5$ $\theta(t) = \theta(0) \pm \pi f_d t = \theta(0) \pm \pi t / 2T_b$ over $0 \le t \le T_b$.

The plus sign corresponds to symbol "1" and the minus sign corresponds to symbol "0". A similar result holds for $\theta(t)$ in the interval $-T_b \le t \le T_b$, though the algebric sign is not necessarily the same in both the intervals. The phase $\theta(0)$ is 0 or π depending on the past history of the modulation process. In the interval $-T_b \le t \le T_b$, the polarity of $\cos(\theta(t))$ depends only on $\theta(0)$ regardless of the sequence of 1's and 0's transmitted before or after $t = 0$. Thus for this time interval, the in-phase component $s_I(t)$ consists of a half-cycle cosine pulse defined as follows:

$$s_I(t) = \sqrt{2E_b / T_b}\cos(\theta(t)) = \sqrt{2E_b / T_b}\cos(\theta(0))\cos(\pi t / 2T_b)$$
$$= \pm\sqrt{2E_b / T_b}\cos(\pi t / 2T_b), \qquad T_b \le t \le T_b \quad (5.115)$$

Where the + sign corresponds to $\theta(0) = 0$ and the minus sign corresponds to $\theta(0) = \pi$. In a similar way, we can find that the quadrature component $s_Q(t)$ over the interval $0 \le t \le 2T_b$ consists of a half-cycle sine pulse, whose polarity depends only on $\theta(T_b)$ given as

$$s_Q(t) = \sqrt{2E_b / T_b}\sin(\theta(t)) = \sqrt{2E_b / T_b}\sin((\theta T_b))\sin(\pi t / 2T_b)$$
$$= \pm\sqrt{2E_b / T_b}\sin(\pi t / 2T_b), \qquad 0 \le t \le 2T_b \quad (5.116)$$

Where the plus sign corresponds to $\theta(T_b) = \pi/2$ and minus sign corresponds to $\theta(T_b) = -\pi/2$. Analysis of the above things provides four possibilities:

1. The phase $\theta(0) = 0$ and $\theta(T_b) = \pi/2$, corresponds to transmission of symbol "1"
2. The phase $\theta(0) = \pi$ and $\theta(T_b) = \pi/2$, corresponds to transmission of symbol "0"
3. The phase $\theta(0) = \pi$ and $\theta(T_b) = -\pi/2$, (equivalently $3\pi/2$ modulo 2π) corresponds to transmission of symbol "1"
4. The phase $\theta(0) = 0$ and $\theta(T_b) = -\pi/2$, corresponds to transmission of symbol "0"

Thus MSK signal can take any one of four possible forms depending on the values of $\theta(0)$ and $\theta(T_b)$.

In terms of the orthonormal basis functions $\varphi_1(t)$ and $\varphi_2(t)$, MSK are defined by a pair of sinusoidally modulated quadrature carriers as

$$\varphi_1(t) = \sqrt{2/T_b}\ \cos(\pi t/2T_b)\cos(2\pi f_c t), \qquad 0 \le t \le T_b \tag{5.117a}$$

$$\varphi_2(t) = \sqrt{2/T_b}\ \sin(\pi t/2T_b)\sin(2\pi f_c t), \qquad 0 \le t \le 2T_b \tag{5.117b}$$

$$s(t) = s_1\,\varphi_1(t) + s_2(t)\,\varphi_2(t), \qquad 0 \le t \le T_b \tag{5.117c}$$

The coefficients s_1 and s_2 are related to the phase states $\theta(0)$ and $\theta(T_b)$ respectively and are given by Eq. (5.118).

$$s_1 = \int_{-T_b}^{T_b} s(t)\,\varphi_1(t)\,dt$$

$$= \sqrt{E_b}\cos(\theta(0)) \qquad -T_b \le t \le T_b \tag{5.118a}$$

$$s_2 = \int_{0}^{2T_b} s(t)\,\varphi_2(t)\,dt$$

$$= -\sqrt{E_b}\sin(\theta(T_b)) \quad 0 \le t \le 2T_b \tag{5.118b}$$

Both the integrals in Eqs. (5.118a) and (5.118b) are evaluated over $2T_b$. The time interval $0 \le t \le T_b$, for which the phase $\theta(0)$ and $\theta(T_b)$ are defined, is common to both integrals. Fig. 5.45 shows the different MSK waveforms.

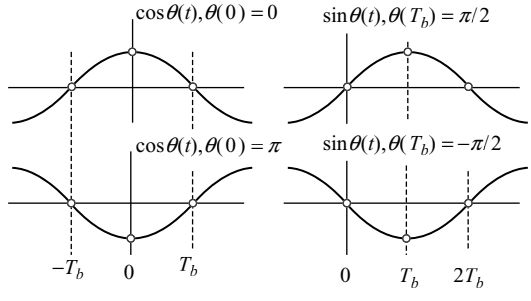


Fig. 5.45 *Different forms of MSK waves*

5.20.1 Signal Constellation of MSK Waveforms

The signal constellation for an MSK signal is two-dimensional ($N = 2$) with four possible message points ($M = 4$). The coordinates of the four message points are:

$$(+\sqrt{E_b}, +\sqrt{E_b})\ (-\sqrt{E_b}, +\sqrt{E_b})\ (-\sqrt{E_b}, -\sqrt{E_b})\ (+\sqrt{E_b}, -\sqrt{E_b})$$

The possible values of $\theta(0)$ and $\theta(T_b)$ corresponding to these message points are also shown in Fig. 5.46. MSK signal-point diagram is similar to QPSK with a subtle difference to be noticed carefully. In QPSK the transmitted symbol is represented by any one of the four message points, whereas in MSK one of two message points is used to represent the transmitted symbol at any one time, depending on the value of $\theta(0)$.

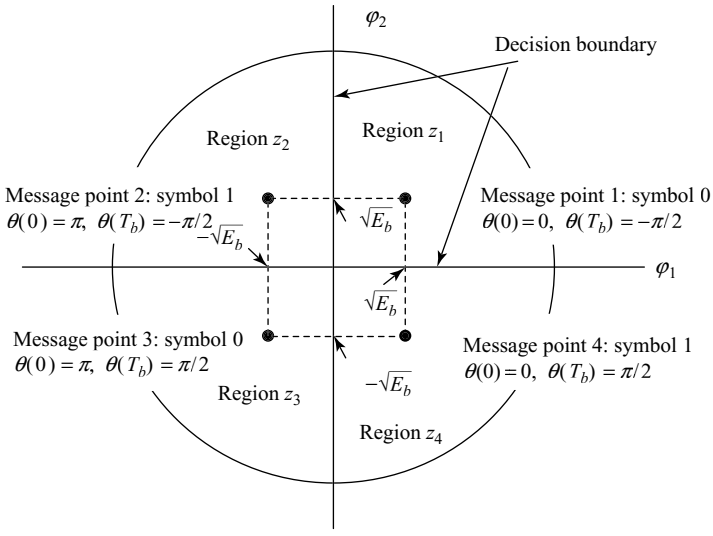


Fig. 5.46 *Signal Space diagram for MSK*

Table 5.3 summarizes the values for s_1 and s_2 for different values of $\theta(0)$ and $\theta(T_b)$ for the interval $-T_b \le t \le T_b$ and $0 \le t \le 2T_b$. From Table 5.3, it is observed that first two symbols (symbol 1) involve up conversion and last two (symbols 0) involves down conversion. Again when the message points have opposite sign symbol 1 is transmitted and has same sign when symbol 0 is transmitted.

Table 5.3 *Signal space characteristics of MSK*

Transmitted binary symbol, $0 \le t \le T_b$	Phase value in radians $\theta(0)$	$\theta(T_b)$	Message points coordinates s_1	s_2
0	0	$-\pi/2$	$+\sqrt{E_b}$	$+\sqrt{E_b}$
1	π	$-\pi/2(3\pi/2)$	$-\sqrt{E_b}$	$+\sqrt{E_b}$
0	π	$+\pi/2$	$-\sqrt{E_b}$	$-\sqrt{E_b}$
1	0	$+\pi/2$	$+\sqrt{E_b}$	$-\sqrt{E_b}$

Let the MSK signal is represented as

$$s_{MSK}(t) = s_I(t) \cos(2\pi f_c t) + s_Q(t) \sin(2\pi f_c t) \tag{5.119}$$

Where the inphase and quadrature components $s_I(t)$ and $s_Q(t)$ are already defined by Eqs. (5.115) and (5.116) respectively.

As described for QPSK signal, the input binary sequence of data bits $d_k = \{0,1\}$ arrives at the modulator input at a rate $1/T$ bits/sec and is separated into two data streams $d_I(t)$ and $d_Q(t)$ containing odd and even bits respectively. Equation (5.119) can be represented in an alternative way as

$$s_{MSK}(t) = d_I(t) \cos(\pi t/2T_b) \cos 2\pi f_c t + d_Q(t) \sin(\pi t/2T_b) \sin 2\pi f_c t$$

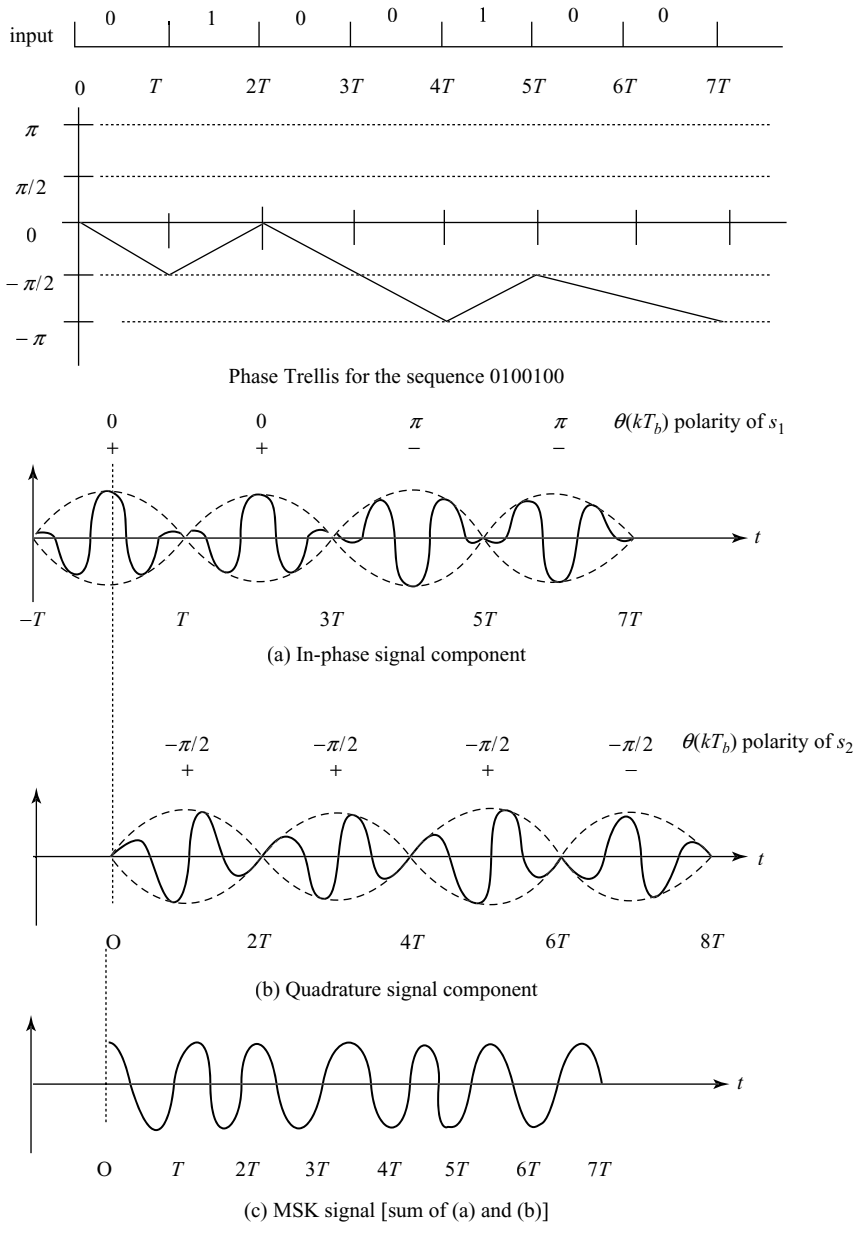
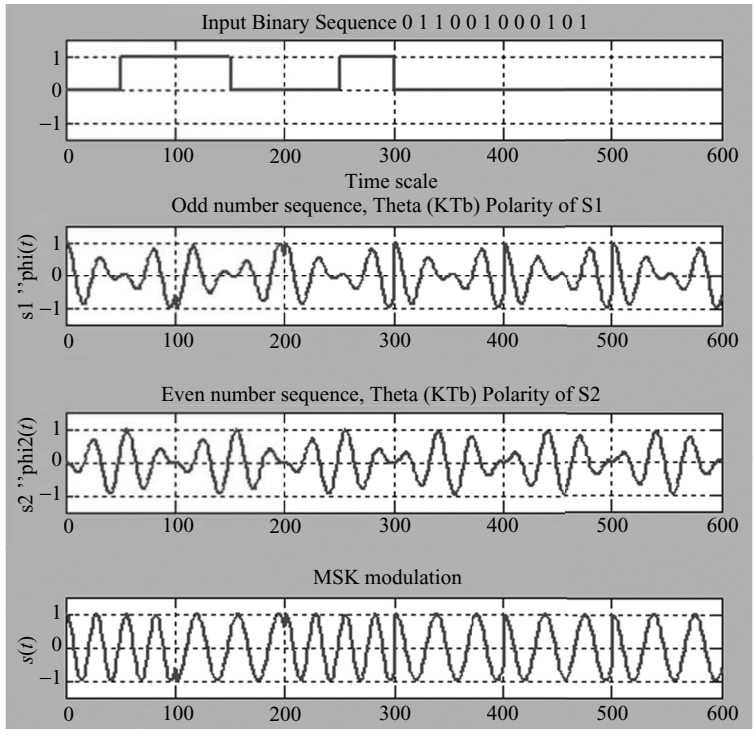
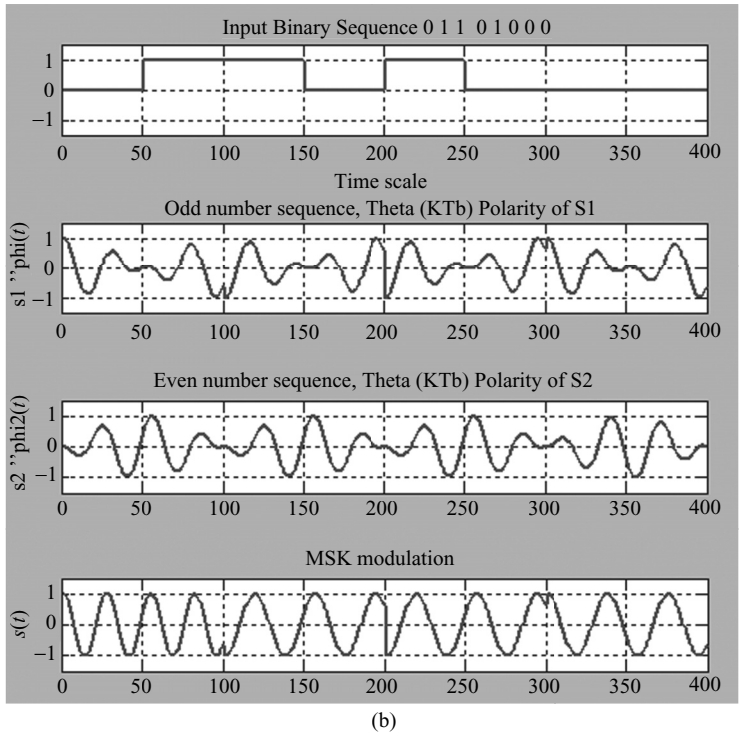


Fig. 5.47 *MSK signal generations*

Figure 5.48 is the generation of the MSK signals using MATLAB simulation for binary bit sequences (a) 011001000101, and (b) 01101000 showing the waveforms of inphase and quadrature components generated from odd number and even number bit sequences respectively. The polarily of inphase and quadrature components is determined from phase trellis for the bit sequences as described in Fig. 5.47.



(a)



(b)

Fig. 5.48 *Generations of MSK waveforms using MATLAB simulator*

5.20.2 Error Probability of MSK Signal

In QPSK the transmitted signal is represented by any one of four messages, whereas in MSK one of two message points is used to represents transmitted signal at any time. The distance between two signal points in MSK is $2\sqrt{E_b}$ and is same as of QPSK.

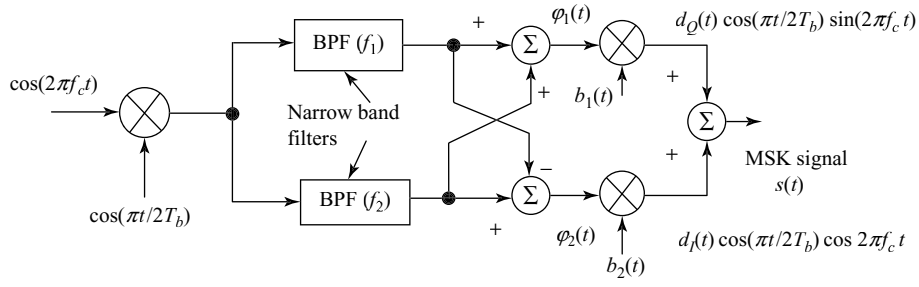
Referring to the signal-space diagram of Fig. 5.46, it is observed that receiver makes decision in between the message points m_1 and m_3 for symbol 1 and in between m_2 and m_4 for symbol 0 according to the estimation of $\theta(0)$ either 0 or π and $\theta(T_b)$ is $-\pi/2$ or $\pi/2$.

Bit decisions are made alternately in the I (in-phase) and Q (quadrature) channels of the receiver over the period $2T_b$ seconds. The receiver makes an error whenever a wrong value of $\theta(0)$ in the I-channel or the wrong value of $\theta(T_b)$ in the Q-channel is obtained. The signal from other bits does not interfere with the receiver's decision for a given bit in a given channel. So, the bit error probability of MSK is exactly same as of QPSK or binary PSK signal and is given by

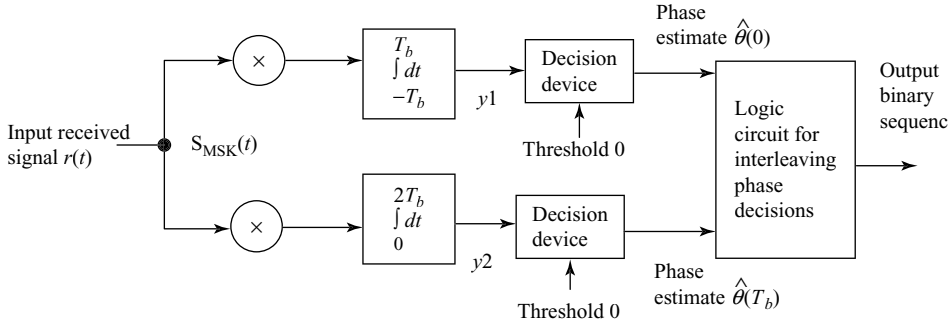
$$P_{e\,\mathrm{MSK}} = 1/2\ \mathrm{erfc}\left[\sqrt{E_b/N_0}\right] \tag{5.120}$$

5.20.3 Generation and Detection of MSK Signals

At the input of MSK modulator of Fig. 5.49(a), there are two sinusoidal waves one with carrier frequency $f_c = n/4T_b$, n is an integer, and other with frequency $f = 1/4T_b$ those are applied at the product modulator. This produces two phase-coherent sinusoidal waves at frequencies f_1 and f_2 related with f_c and T_b as $f_c = (f_1 + f_2)/2$, and $h = T_b(f_1 - f_2) = 0.5$. Two narrowband filters are used to separate these frequencies. The resulting filter outputs are combined to produce two orthonormal signals $\varphi_1(t)$ and $\varphi_2(t)$. Multiplying $\varphi_1(t)$ and $\varphi_2(t)$ by two binary waves $b_1(t)$ and $b_2(t)$ with bit rate $1/2T_b$ and added up to produce MSK signal. These binary waves are extracted from the information bit stream by using serial to parallel converter.



(a) MSK modulator



(b) MSK demodulator

Fig. 5.49 *MSK generation and detection circuits*

For demodulation in Fig. 5.49(b), the received signal is multiplied with coherently generated functions $\varphi_1(t)$ and $\varphi_2(t)$. Integrating in the two separate in-phase and quadrature channels that produce correlation output y_1 and y_2 to be compared with the threshold 0 in the decision devices. The estimation is made for $\theta(0)$ in the I-channel and for $\theta(T_b)$ in the Q-channel. Finally, the phase decisions are interleaved so as to reconstruct the original input binary sequence with the minimum average probability of symbol error in an AWGN channel.

5.20.4 Power spectra of MSK Signals

The input binary wave is random in transmission of bits 1 and 0, i.e., equally likely. The symbols transmitted during different time slots being statistically independent. The inphase and quadrature components of the MSK signal is given by Eqs. (5.115) and (5.116) as given under for recapitulation.

$$s_1(t) = \sqrt{2\ E_b\ /T_b}\ \cos\ (\theta(t)) = \sqrt{2\ E_b\ /T_b}\ \cos\ (\theta(0))\cos\ (\pi t\ /2T_b)$$
$$= \pm\sqrt{2\ E_b\ /T_b}\ \cos\ (\pi t\ /2T_b), \quad -T_b \le t \le T_b$$
$$s_Q(t) = \sqrt{2\ E_b\ /T_b}\ \sin\ (\theta(t)) = \sqrt{2\ E_b\ /T_b}\ \sin\ (\theta(T_b))\sin\ (\pi t\ /2T_b)$$
$$= \pm\sqrt{2\ E_b\ /T_b}\ \sin\ (\pi t\ /2T_b), \quad 0 \le t \le 2T_b$$

Depending on the value of $\theta(0) = 0$, the in-phase component in terms of pulse shaping filter is + g(t) and for $\theta(0) = \pi$ it is $-g(t)$. In a similar way, we can find that the quadrature component $s_Q(t)$ over the interval $0 \le t \le 2T_b$ is + $g(t)$ or $-g(t)$ depending on phase state value $\theta(T_b)$ $+\pi/2$ and $-\pi/2$ respectively.

Thus,
$$g(t) = \begin{cases} \sqrt{2E_b\ /T_b}\ \cos\ (\pi t\ /2T_b), & -T_b \le t \le T_b \text{ for in-phase component} \\ 0, \text{ otherwise} \end{cases} \tag{5.121}$$

$$g(t) = \begin{cases} \sqrt{2E_b\ /T_b}\ \sin\ (\pi t\ /2T_b), & 0 \le t \le 2T_b \text{ for quadrature component} \\ 0, \text{ otherwise} \end{cases} \tag{5.122}$$

The energy spectral density for both of these components is equal to (obtained by Fourier transform),

$$\Psi_g(f) = (32\ E_bT_b)\ /\pi^2\ [\cos\ (2\pi fT_b)\ /\ (16\ T_b^2\ f^2\ -1)\ ]^2 \tag{5.123}$$

So, the power spectral density of the MSK signal is

$$S_{\text{MSK}}(f) = 2\ \text{x}\ \Psi_g\ (\text{f})\ /\ 2\ T_b$$
$$= (32E_b\ /\pi^2)\ [\ \cos\ (2\pi fT_b)\ /\ (16\ T_b^2\ f^2\ -1)\ ]^2 \tag{5.124}$$

The MSK modulation makes the phase change linear and limited to $\pm\ (\pi/2)$ over a bit interval T_b. This enables MSK to provide a significant improvement over QPSK. As shown in Fig. 5.50, the MSK signal falls off as the inverse fourth power of frequency, whereas QPSK falls off as the inverse square of frequency. Because of the effect of the linear phase change, the power spectral density has low side lobes that help to control adjacent-channel interference. For a band limited digital communication system this is very advantageous. However, the main lobe becomes wider than the quadrature shift keying.

But the fundamental problem with MSK is that the spectrum has side-lobes extending well above the data rate. Thus for wireless systems that require more efficient use of RF channel Bandwidth, it is necessary to reduce the energy of the upper side-lobes. As a solution to this problem a pre-modulation filter can be used (Low pass filter) or the more efficient and real approach is the use of Gaussian Filter that necessitates the use of GMSK in GSM wireless networks. Proper utilization of phase during detection is made in MSK for improving noise performance.

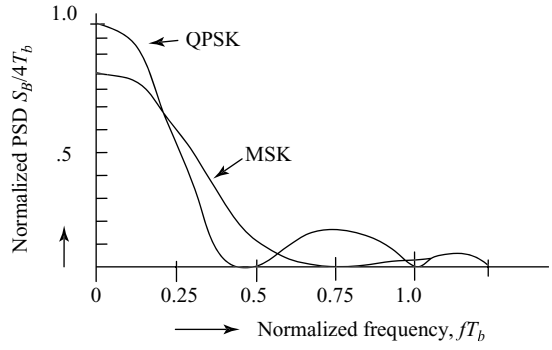


Fig. 5.50 *Normalized power spectral density (PSD) plot with normalized frequency for MSK and QPSK*

5.21 GAUSSIAN MINIMUM SHIFT KEYING: GMSK

In MSK, we replace the rectangular pulse with a sinusoidal pulse. Obviously, other pulse shapes are possible. A Gaussian-shaped impulse response filter generates a signal with low side lobes and narrower main lobe than the rectangular pulse. Since the filter theoretically has output before input, it can only be approximated by a delayed and shaped impulse response that has a Gaussian—like shape. This modulation is called Gaussian Minimum Shift Keying (GMSK).

Before going into the GMSK, let us discuss about the pulse shaping of signals. In general, the square pulse is not useful in sending information, as it requires very large bandwidth. Further generating perfect square pulse is difficult. In lieu of pulse, shaped pulses that convey the same information but use smaller bandwidths and having good properties like inter-symbol interference rejection are used. The very popular pulse shaping is called *root raised cosine* that has a parameter called roll-off which controls the shape and the bandwidth of the signal. Some very common pulse shaping methods are:

1. Root raised cosine used in QPSK
2. Half-sinusoid used in MSK and
3. Gaussian pulse used in GMSK

The frequency response function for root raised cosine is defined as

$$P(f) = \begin{cases} 1/2w, & 0 \le |f| < f_1 \\ 1/4w\left\{1-\sin\left[\pi(|f|-w)/(2w-2f_1)\right]\right\}, & f_1 \le |f| \angle 2w - f_1 \\ 0 & |f| \ge 2w - f_1 \end{cases} \tag{5.125}$$

The frequency f_1 and bandwidth w is related to a parameter α known as the roll-off parameter as described

$$\alpha = 1 - f_1/w \tag{5.126}$$

It indicated the excess bandwidth over the ideal solution, w. Specifically, the transmission bandwidth B_T is defined by

$$B_T = 2w - f_1 = w\ (1 + \alpha)$$

To see the relation between α and frequency response of $P(f)$, normalized frequency response is plotted for different values of α in Fig. 5.51.

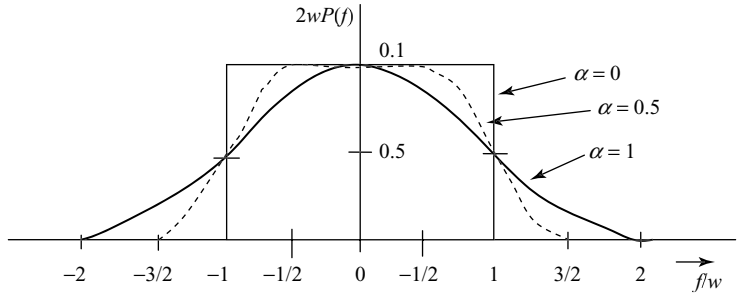


Fig. 5.51 *Frequency response of signal with different roll-off factors*

$\alpha = 0$ corresponds to ideal Nyquist channel. The time response of the shaping pulse is the inverse Fourier transform of $P(f)$.

$$p(t) = \{\text{sinc}(2wt)\} \text{ x } [\cos(2\pi\alpha wt) / (1 - (4\alpha wt)^2)] \quad (5.127)$$

The time response $p(t)$ consists of two terms. One is the sinc function that represents the ideal Nyquist channel, and ensures zero crossing of $p(t)$ at the desired sampling instant $t = nT_s$, n is the integer (positive or negative), T_s is sampling rate. The second factor reduces the tails of the pulse considerably below the value that obtained from the ideal Nyquist channel. For $\alpha = 1$, we see the most gradual rolloff in the amplitude of the oscillatory tails of $p(t)$. Thus the amount of intersymbol interference resulting from timing error decreases as the roll-off factor α is increased from zero to unity.
Considering $\alpha = 1$, Eq. (5.127) reduces to

$$p(t) = \text{sinc}(4wt) / (1 - 16\, w^2 t^2) \quad (5.128)$$

Figure 5.52 is the time response $p(t)$ with different roll-off factors.

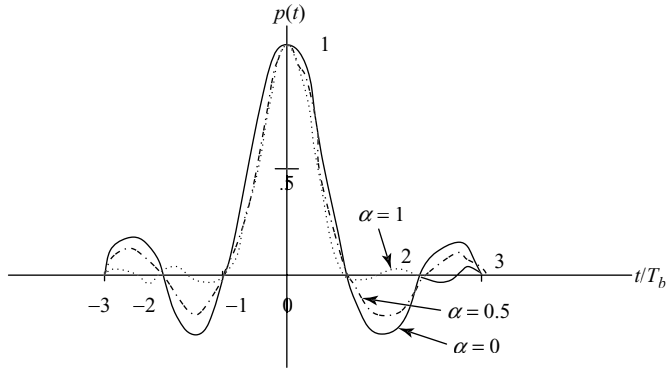


Fig. 5.52 *Time response p(t) for different roll-off factors*

This special case of $\alpha = 1$ ($f_1 = 0$) is called the full-cosine roll-off characteristics, for which the frequency response $P(f)$ is defined as follows:

$$P(f) = (1/4w)\left[1 + \cos(\pi f / 2w)\right],\ 0 < |f| < 2w$$
$$= 0,\ |f| > 2w \quad (5.129)$$

Shaping the spectrum should satisfy two criteria: The main lobe should be as narrow as possible, and the maximum side lobe level should be as small as possible relative to the main lobe.

GMSK modulation is based on MSK, a phase shift keying technique. The basic problem of PSK is the extension of sideband from the carrier. MSK and GMSK are used to overcome this problem, which are continuous phase shift techniques as the frequency changes occur at carrier zero crossing points. However, the out-of-band spectral density of the MSK signal does not satisfy the stringent requirement in wireless application. GMSK uses spectrum efficiently and most widely used in GSM cellular technology. To reduce the spectrum extension into the side band, in MSK Pulse Shaping filter is applied prior to applying in the carrier. The design of filter will be such that it has narrow bandwidth, sharp cut-off and impulse response should have no overshoot. The most important point of GMSK is that it uses Gaussian filter whose spectral response to an impulse is also Gaussian without ringing. The relationship between the pre-modulation filter bandwidth B and the bit period T defines the bandwidth of the system. GSM designers used a BT = 0.3 with a channel data rate of 270.8 kbps. This compromises between a bit error rate and an out-of-band interference since the narrow filter increases the intersymbol interference and reduces the signal power.

The Gaussian low-pass filter has an impulse response given by the following equation:

$$g(t) = \frac{1}{2T_b}\left[Q\left(2\pi B_g \frac{t - T_b/2}{\sqrt{\ln 2}} \right) - Q\left(2\pi B_g \frac{t - T_b/2}{\sqrt{\ln 2}} \right) \right], 0 \le B_g T_b \le \infty \tag{5.130}$$

and Q(t) is the Q-function defined as

$$Q(t) = \int_t 1/\sqrt{2}\ \exp\ (-x^2/2\ dx)$$

B_g is the bandwidth of the low pass filter having a Gaussian shaped spectrum, T_b is the bit period and $B_N = B_g T_b$ is the normalized bandwidth. To demonstrate this, we are looking at a filter with a bandwidth of $B_g = 1000$ and a bit rate of $T_b = 1/2000$, i.e., a normalized bandwidth $B_N = B_g.T_b = 0.5$. The impulse response of the Gaussian low-pass filter has to be truncated and scaled, according to the B_N value, to ensure that the effect of a single 1 passing through the filter has a phase change of $\pi/2$. For $B_N = 0.5$ the filter response is truncated, symmetrically around zero, to two bit periods from $-T$ to T. The truncated filter response is represented graphically in the following Fig. 5.53.

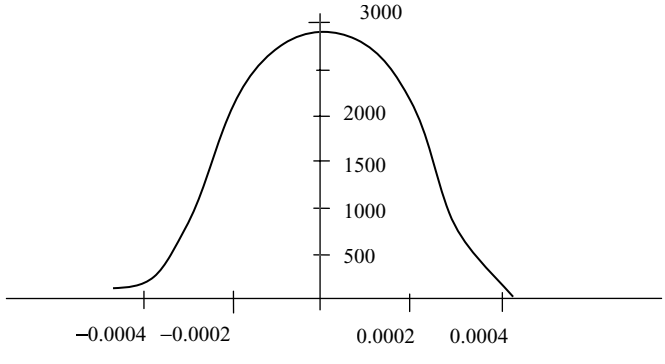


Fig. 5.53 *Truncated filter response of Gaussian pulse*

In order to reduce the sidelobes and to produce a compact spectrum, the $B_g T_b$ factor controls the effects of the Gaussian filter. Generally, B_g is the 3db (or half-power) bandwidth.

Consider W_{3dB}, a factor refers to the filter's –3dB bandwidth and data rate defined by: $W_{3dB} = f_{\text{-}3dB}$ / Bit rate (T_b). Thus $W_{3dB} T_b$ is the time bandwidth product, an important parameter for GMSK modulation. When this product tends to infinity then it corresponds to the normal MSK operation, and when less than 1,

then more power will concentrate inside the pass band of GMSK signal. A typical relationship between the Spectral Density with different time-bandwidth product is shown in Fig. 5.54.

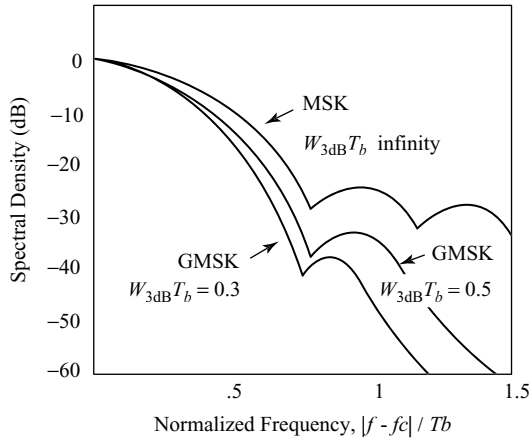


Fig. 5.54 *Power spectral density with different time-bandwidth product in MSK and GMSK signal*

One of the advantages of GMSK modulation is the spectral efficiency compared to other PSK modulation. Other is that the GMSK can use non-linear amplifier and remain undistorted. This helps to use in small portable amplifiers.

GSM is an ideal choice for many applications. An important application of GMSK is in GSM system, which is a time-division multiple-access system. For this application, the $W_{3dB}T_b$ is standardized at 0.3, which provides the best compromise between increased bandwidth occupancy and resistance to Inter Symbol Interference (ISI). Ninety-nine percent of the RF power of GMSK signals is specified to confine to 250 kHz, which means that the sidelobes need to be virtually zero outside this frequency band and the ISI should be negligible.

5.22 ORTHOGONAL FREQUENCY DIVISION MULTIPLEXING: OFDM

Digital multimedia applications are getting ever-increasing demand for broadband communication systems. OFDM is a method that allows transmitting high data rates over wireless noisy channels (delay dispersive environments) at a comparable low complexity. In OFDM the entire frequency bands is divided into number of subcarriers. Spread spectrum technique is used to distribute data over these sub carriers that are separated with orthogonality conditions. The benefits of OFDM are high spectral efficiency because of parallel data transmission. It is the modulation technique used for digital TV in Europe, Japan and Australia.

5.22.1 Concept of Parallel Transmission: Single Carrier vs. Multicarrier

The illustration given in Fig. 5.55 is self-explanatory for parallel transmission scheme for OFDM system. Fig. 5.55(a) is the transmission technique used in code division multiple access (multicode), and (b) is the technique for multi-carriers where N number of subcarriers are used.

Now question arises why multicarrier than single carrier. The ISI is more in single carrier system as the channel is frequency selective, whereas ISI is greatly reduced in case of multicarrier transmission as flat fading occurs. Higher delay spread leads to higher ISI. Multimedia transmission requires high data rate, so symbol duration is very small and signal is more and more affected by ISI when data rate increases. As multicarrier transmission divides the available carrier bandwidth into smaller sub-bands, represented by subcarriers whose data rate is smaller. So, symbol duration at the subcarrier increases and automatically ISI decreases.

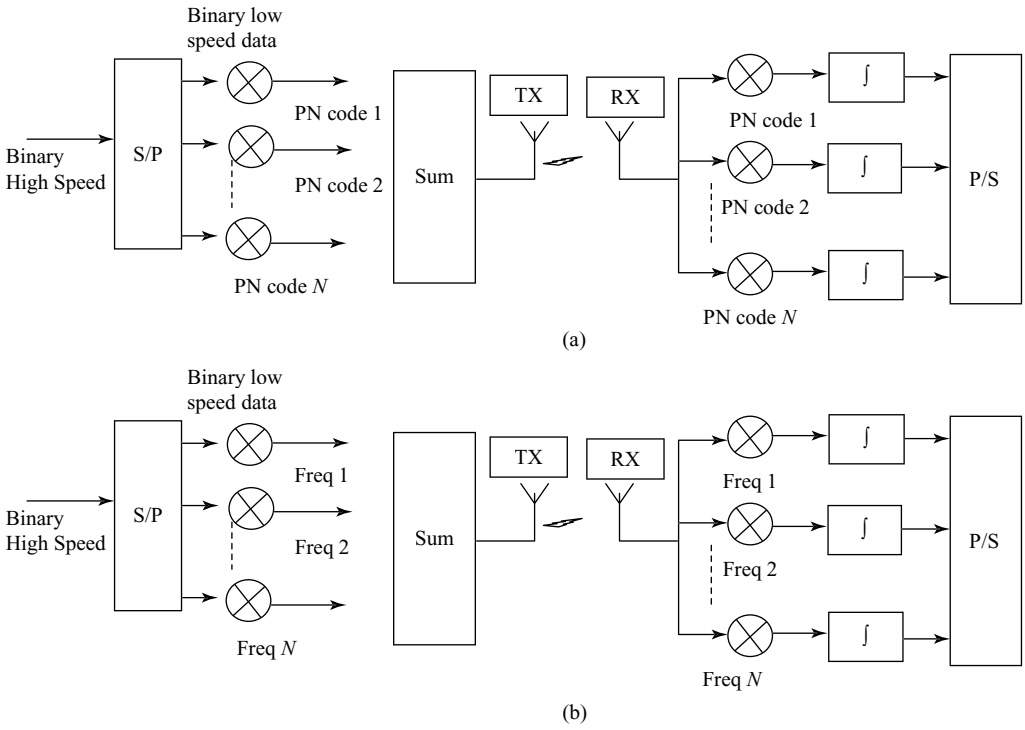


Fig. 5.55 *(a) Multicode, (b) Multicarrier transmission techniques*

As an example, suppose the transmission data rate in a single carrier system is $R = 1/T = 8$ *M* symbol /sec, the maximum channel delay $\tau_{max} = 10$ μsec.

The ISI = $\tau_{max}/T = 10 \text{ x } 10^{-6} \times 8 \times 10^{6} = 80$

For multicarrier, ISI = $\tau_{max}/N \times T$, *N* is the number of subcarriers (say 256). Then ISI = 80/256 = 0.3 which is very low.

Multicarrier has the following characteristics:

1. Divides the bit stream into N substreams
2. Modulates substream with bandwidth *W*/*N*
3. Separate subcarriers
4. $B/N < B_c$ flat fading (no OFDM ISI)

5.22.2 OFDM Basics

OFDM – Orthogonal FDM. *N* subchannel signals generated jointly to make sure that they are orthogonal to each other. In traditional FDM, signals are generated separately for each subcarriers. So, in case of frequency resource allocation, OFDM is very similar to conventional FDM. The difference of OFDM with FDM is that signals are orthogonal to each other in OFDM, while they are not orthogonal in FDM. This is because, signals are generated together and they can be made 'orthogonal'. This orthogonality is a very important characteristic of OFDM systems. Orthogonal signal can be transmitted via overlapped spectrum, so significant amount of bandwidth can be saved. This is simply illustrated in Fig. 5.56. Where $N = 8$ number of channels are used in FDM system with guard bands in between channels. In case of OFDM due to orthogonality, signals are overlapped and can save up to 50% of bandwidth. Data are shared among several carriers and simultaneously transmitted in OFDM.

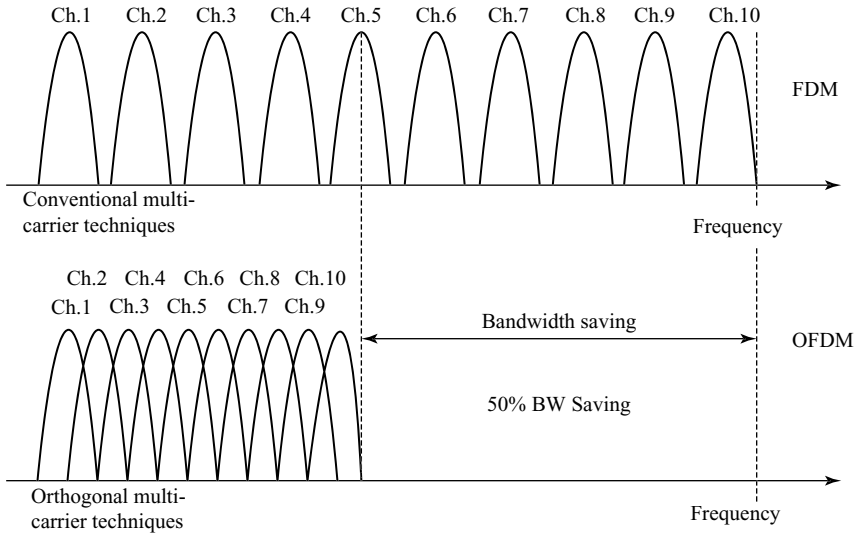


Fig. 5.56 *OFDM modulation showing 50% BW saving*

Figure 5.57 is the representation of time-frequency relationship of the OFDM signal. Inter carrier separation = Any integer multiple of 1/(symbol duration). The distinct features of OFDM are:

1. No carrier guards bands
2. Controlled overlapping of bands
3. Maximum spectral efficiency (Nyquist rate)
4. Easy implementation using FFTs
5. Very sensitive time-freq synchronization

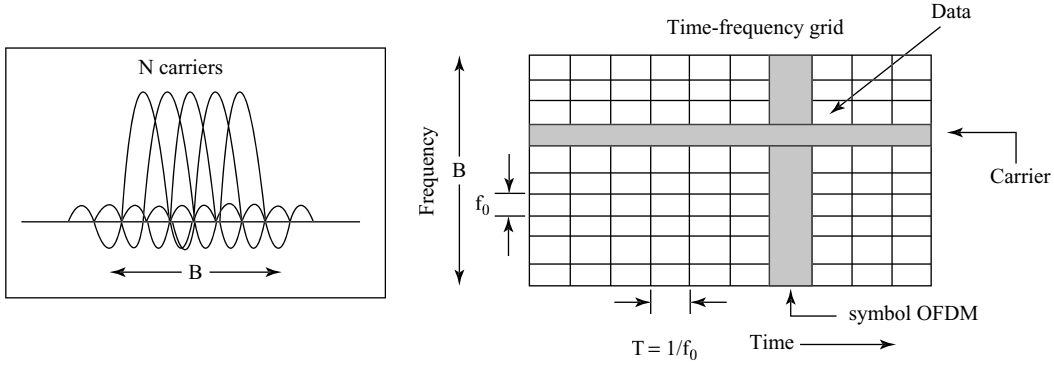


Fig. 5.57 *Time–frequency relationship of OFDM signal*

In OFDM modulation, one user utilizes all carriers simultaneously to transmit its data whereas in the access technique of FDMA several users share dynamically the carriers (traffic or service dependent) to access to the system.

The advantages of OFDM are:

1. Flat fading per carrier
2. N long pulses
3. ISI is comparatively low
4. Easy to exploit frequency diversity
5. Use of dynamic signaling and adaptive coding

In OFDM, the subcarrier pulse used for transmission is considered as rectangular to get the advantage of pulse forming and modulation to be performed by a simple **Inverse Discrete Fourier Transform** (IDFT) that can be implemented very efficiently by **Inverse Fast Fourier Transform** (IFFT). At the receiver only it needs a FFT to reverse this operation. Fourier transform of the rectangular pulse has the form of $\sin(x)/x$ type of spectrum of the subcarriers as shown in Fig. 5.58.

Obviously the spectrums of the subcarriers are not separated but overlaped. Orthogonal signals can be transmitted via overlapped spectrum. So, significant amount of bandwidth is saved. The signals can still be well recovered in the receiver, provided that the orthogonality is maintained.

The multicarrier transmission divides the available carrier bandwidth into smaller sub-bands called sub-carriers. OFDM can achieve large delay spread tolerance at high bit rates by converting single bit stream into N parallel bit streams. Each parallel bit stream is modulated on one of N subcarriers.

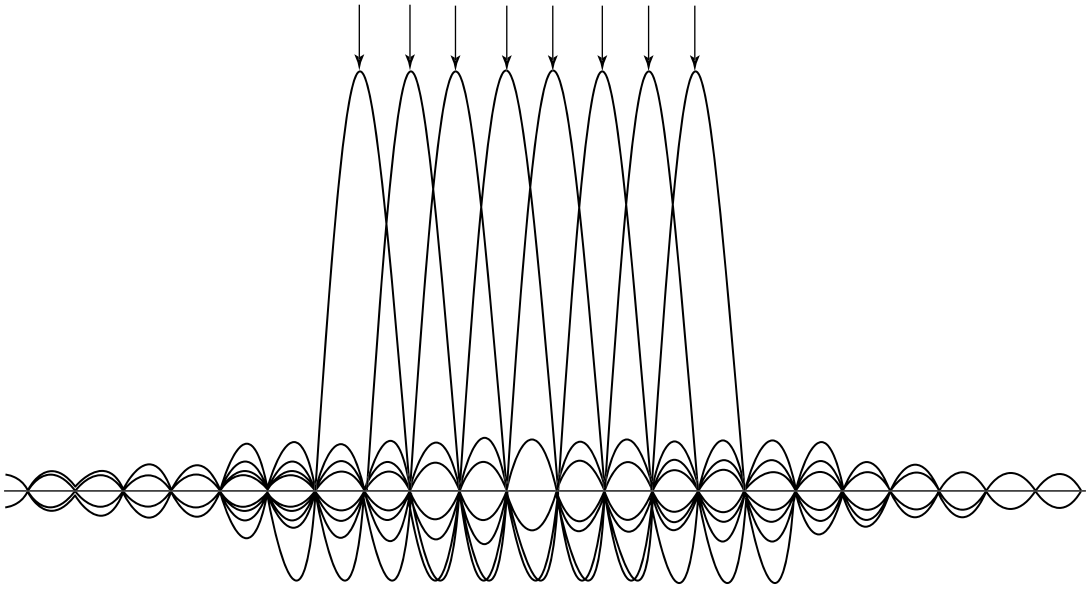


Fig. 5.58 *OFDM spectrums*

Subcarrier data rate is smaller, and the symbol duration is increased. So, relative delay spread decreases, it automatically reduces the ISI. Since the duration of each symbol is long, it is feasible to insert a guard interval between the OFDM symbols, thus eliminating the intersymbol interference. The guard loss can be reduced by choosing N (subcarriers) large enough. In multipath environment, fading introduces bit error. Error coding is used for subcarriers and for improved performance average received power is considered rather than the lowest subcarrier power.

Two signals are orthogonal two each other if they satisfy the following condition:

$$\int_0^T \psi_k\,(t)\,\psi_i^*(t)\,dt = \begin{cases} 1, & \text{if } i = k, \\ 0, & \text{if } i \neq k \end{cases} \tag{5.131}$$

The mathematical expresion for a single carrier signal is represented as

$$s_c\,(kT) = A_c\,(t)\,e^{j(\omega_c\,kT + \phi_c)} \tag{5.132}$$

The multiple orthogonal carriers is mathematically expressed as

$$S_m(kT) = \frac{1}{N}\sum_{n=0}^{N-1} A_n\,e^{j\,\varphi_n}\,e^{j\,(\omega_0\,+\,n\,\Delta\omega)\,kT} \tag{5.133}$$

Sampled in time domain — Frequency domain representation

In inverse Fourier transform, this is equivalent to

$$g(kT) = \frac{1}{N}\sum_{n=0}^{N-1} \mathrm{G}(n/NT)\,e^{j\,2\pi\,nk/N} \tag{5.134}$$

5.22.3 Baseband Analytical OFDM Model

Assume the information data sequence be $X(k)$, $k = 0,1, \ldots, N-1$, then the transmited signal $s(t)$ can be expressed as

$$s(t) = \sum_{n=0}^{N-1} X(p)\exp\,(j\,2\pi f_p t)$$

$$= \sum_{n=0}^{N-1} X(p)\exp\,(j\,2\pi p\Delta f t),\ \ 0 \le t \le T;\ \ 0 \le p \angle N-1$$

The received signal, $r(t) = s(t) * h(t) + w(t)$ (5.136)

where h(t) is the impulse response of the channel and w(t) is the AWGN. The received signal on subcarrier k_0 is

$$Y(k) = (1/T)\int_0^T r(t)\exp\,(-j2\pi k\Delta f t)\,dt \tag{5.137}$$

Assuming $h(t) = 1$ and ideal channel with $w(t) = 0$, $Y(k)$ becomes

$$Y(k) = (1/T)\int_0^T \sum X(p)\exp\,(j2\pi p\Delta f t)\exp\,(-j2\pi k\Delta f t)$$

Transmitted data can be fully recovered from above Equation if $p = k$ in $Y(K)$.

5.22.4 Modulation and Demodulation of OFDM Signal Using Analog Technique

Figure 5.59 is the illustration for OFDM modulation and demodulation techniques. The binary input data is first encoded and then converted into N parrallel bit streams. It is then modulated with N numbers of carriers with frequency f_k, $k = 0$ to N-1. The output of the productor is summed up to get OFDM signal which is passed throug a Low Pass Filter (LPF) and then transmitted over wireless channel with channel response $h(t)$.

At the receiver the received signal is passed to the N productor units where it is multiplied by the same subcarrier signals. The output of the productors are then integrated over 0 to T to retrieve the estimated signals $\tilde{A}_{k,0}$ …. $\tilde{A}_{k,N-1}$ which are then converted from parallel to serial data, after decoding the binary form of transmitted signal is obtained.

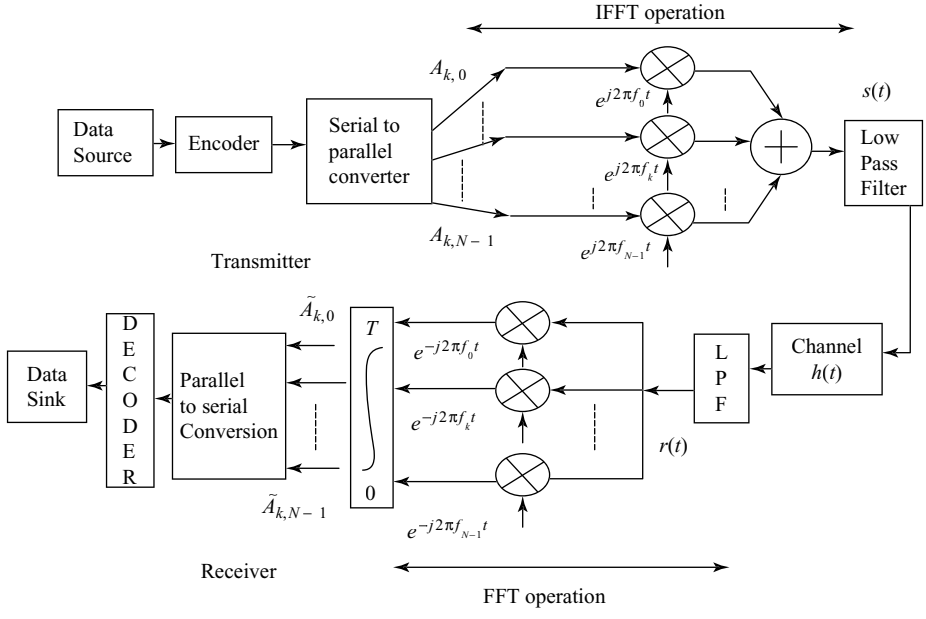


Fig. 5.59 *OFDM modulation and demodulation by analog technique*

5.22.5 OFDM Modulation Unig FFTS

Let us assume that $\psi_k(t) = p(t)\, e^{j2\pi f_k t}$; where $f_k = k/T$, so the signal $s(t)$ can be expressed as

$$s(t) = \sum_{k=0}^{N-1} \sum_{m} c^k m\ p(t - mT)\, e^{j2\pi f_k t} \tag{5.139}$$

where c is the data and p is the pulse.

The k^{th} carrier modulator is

Sampling at $T_s = T/N$,

$$s(n) = \sum_{k=0}^{N-1} \sum_{m} c^k{}_m\ \text{rect}\,(nT_s - mnT_s)\, e^{(j2\pi kn\ T_s / NT_s)}$$

$$= \sum_{k=0}^{N-1} \sum_{m} c^k{}_m\ \text{rect}\,(n - mN)\, e^{j2\pi kn/N}$$

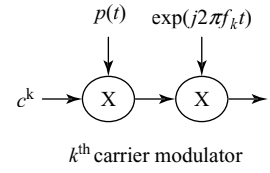


Fig. 5.60 *Carrier modulator unit*

$$= \sum_{m} rect\,[n - mN]\; IDFT\;(c_{m,}\; n)$$

$$IDFT(c_m, n) = \sum_{k=0}^{N-1} c^k m\, e^{j2\pi kn/N}$$

One OFDM data carries N data c, thus $s[n] = \text{IDFT}(c_m, n)$, $\quad mN < n \leq (m+1)\,N$

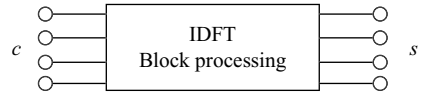


Fig. 5.61 *OFDM data processing using IDFT*

In practice, parallel data modulation and coherent demodulation can be simply done by using IFFT and FFT as shown in Fig. 5.62. The term cyclic prefix in the figure will be discussed later.

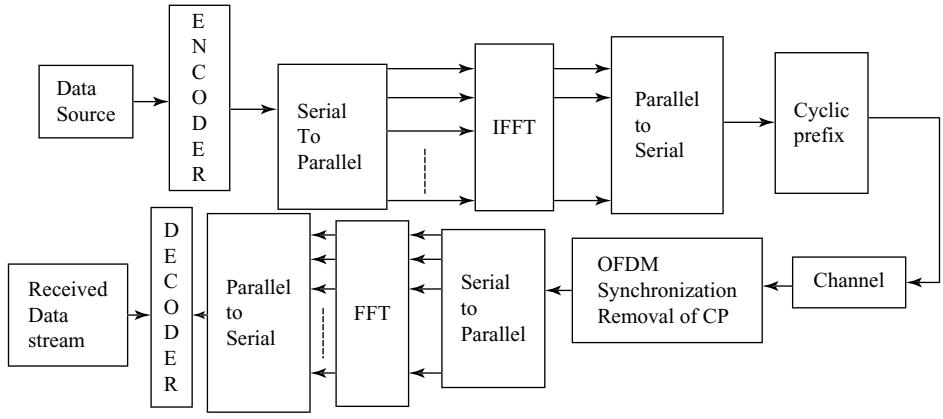


Fig. 5.62 *OFDM transreceiving system using digital technique*

The spacing of the subcarrier can be selected by choosing IFFT modulation. In almost all practical cases, the number of samples N is chosen to be power of 2. The orthogonality requires that the subcarrier spacing is $\Delta f = k/T_s$, where k is an integer and T_s is the symbol duration. $T_s = N/W + T_{cp}$, where W is the total bandwidth, N is the number of subcarriers, T_{cp} the guard interval is inserted prior to the OFDM block. During this interval, a cyclic prefix is transmitted such that the signal in the interval $-T_{cp} \leq t < 0$ is equal to the signal in the interval $(T_s - T_{cp}) \leq t < T_s$. This T_{cp} is useful to combat the delay dispersion due to multipath effects in the channel. This is shown in Fig. 5.63.

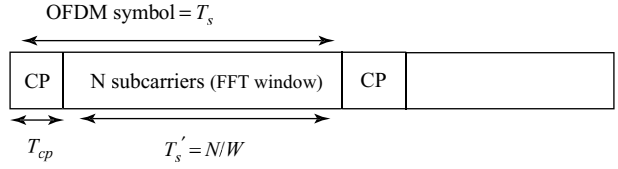


Fig. 5.63 *OFDM symbol duration*

5.22.6 Discrete OFDM Model

CP – is the cyclic prefix, g_l is the channel impulse response padded with zeroes to reach the lengh of N and n_l is uncorrelated Gaussian noise.

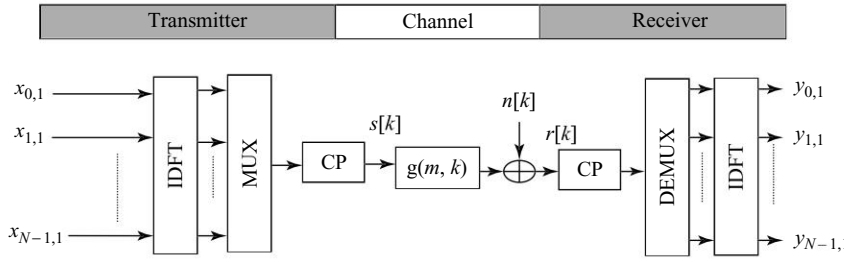


Fig. 5.64 *Discrete OFDM model*

The output signal $y_1 = x_1.h_1 + n_1$, $y_1 = \text{DFT}\ (\text{IDFT}\ (x_1 \otimes g_1) + n_1$ (5.142)
where h_l is the frequency response of the channel = DFT(g_l)

5.22.7 OFDM Guard Interval and Cyclic Prefix

The guard interval or guard time is chosen larger than the expected delay spread occurs due to multipath propagation such that multipath components from one symbol cannot interfere with the next symbol. The guard time could consist of no signal at all. However, in that case the problem of Inter Carrier Interference (ICI) would arise. ICI is cross-talk between different subcarriers, which means they are no longer orthogonal. The effect of guard interval is clearly illustrated in Fig. 5.65.

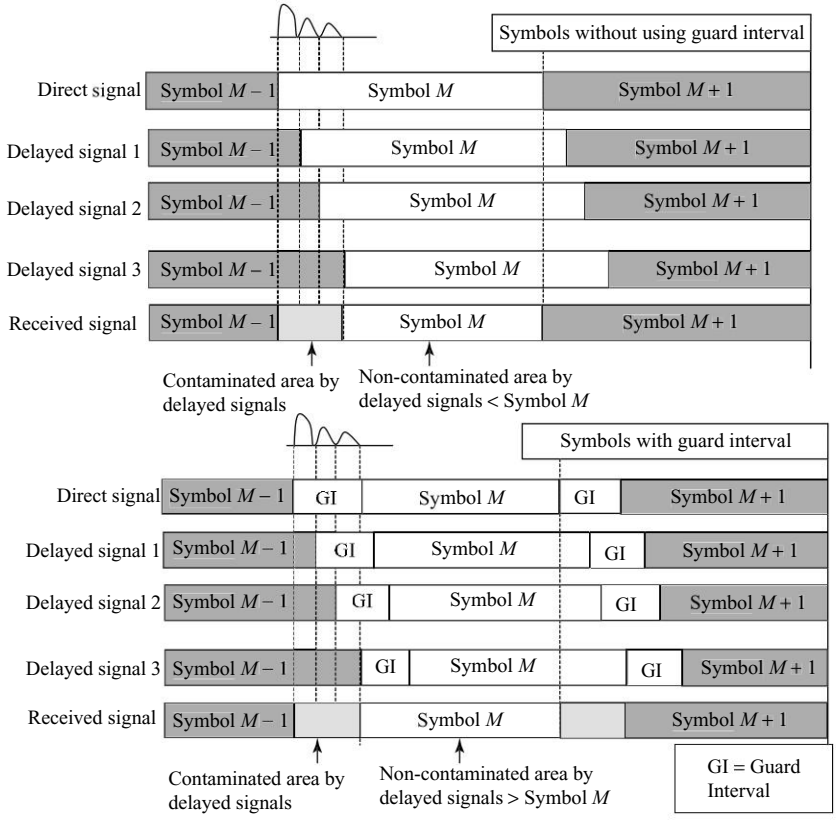


Fig. 5.65 *OFDM symbols without and with guard interval*

Guard Time vs Cyclic Prefix CP is the 'special' time guards in the symbol transitions. Three main functionalities of CP are:

1. It accommodates the decaying transient of the previous symbol (ISI).
2. It avoids the initial transient reaches the current symbol (ICI).
3. CP helps to maintain orthogonality between signals.

This is illustrated in Fig. 5.66.

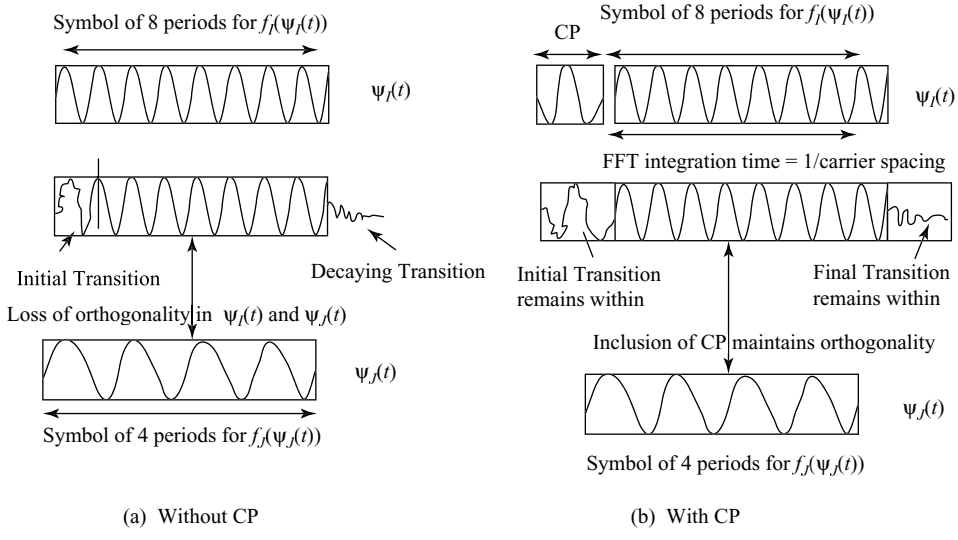


Fig. 5.66 *Effect of cyclic prefix on orthogonality ICI and ISI*

To eliminate ICI, the OFDM symbol is **cyclically extended** in the guard time. This ensures that delayed replicas of the OFDM symbol always have an integer number of cycles within the FFT interval, as long as the delay is smaller than the guard time. As a result, multipath signals with delays smaller than the guard time cannot cause ICI. Multipath delay spread causes ISI, and also loss of orthogonality causes ISI. CP helps the signal to be protected against both ISI and ICI. But CP causes partial loss of signal energy which is given by

$$SNR_{\text{loss-}CP} = -10 \log_{10}(1 - T_{cp} / T_s) \tag{5.143}$$

5.22.8 Peak-to-Average Power Ratio: PAPR

One disadvantage of OFDM is that the peak to the signal can be up to N times the average power, where, N is the number of carriers. OFDM signals have a higher PAPR than single carrier signal. The reason is that in the time domain, a multicarrier signal is the sum of many narrowband signals. At some time instances, this sum is large and at other times is small, which means that the peak value of the signal is substantially larger than the average value. These large peakes increase the amount of intermodulation distortion resulting in an increase in the error rate. The average signal power must be kept low in order to prevent the transmitter amplifier limiting. The non-linear effects on the transmitted OFDM symbols are spectral spreading, intermodulation, and changing the signal constallation. In other words, non-linear effects causes both in-band and out-of-band interference to signals. The in-band interference increses BER of the received signal through warping of the signal constallation and intermodulation, whereas, out-of-band interference causes adjacent channel interference through spectral spreading, the effects limits the usage of OFDM in many systems even they satisfy tolerable in-band interference. Minimizing the PAPR allows a higher average power to be transmitted

for a fixed peak power, improving the overall signal to noise ratio at the receiver. If RF power amplifiers are operated without largr power back-offs, it is possible to keep the out of band power below specified limits. This leads to very inefficient amplification and expensive transmitters. So it is therefore important to minimize the PAPR. A large amount of research has been done on this issues. Selective mapping involves generating a large set of data vectors all representing the same information. The data vector with the lowest resulting PAPR is selected. Information about which particular data vector was used is sent as additional carriers. Another tecnique is the Golay sequences that reduces 3–6 dB PAPR.

Quantifying PAPR In a multicarrier signaling system like OMDM, the resulting waveform is the superposition of N narrowband signals. In particular each of the N output samples form am N-pt IFFT operation involves the N complex numbers. As a cental limit theorem, the resulting output values $\{y_1, y_2, \ldots y_N\}$ can be accurately modeled, particularly for large N, as complex Gaussian random variables with zero mean and variance σ^2, the amplitude of the output signal is

$$| y[n] | = \sqrt{(\text{Re}\{y[n]\})^2 + (\text{Im}\{y[n]\})^2} \tag{5.144}$$

Which is Rayleigh distributed with parameter σ^2. The output power is, therefore,

$$|y[n]|^2 = (\text{Re}\{y[n]\})^2 + (\text{Im}\{y[n]\})^2 \tag{5.145}$$

And the output power is exponentially ditributed with mean $2\sigma^2$. The important thing is that the output amplitude and hence power are random, so the PAPR is not a deterministic quantity. The PAPR is considered for a single OFDM symbol, which consists of $N + N_g$ samples. The discrete time PAPR can be defined for the IFFT output as

$$\text{PAPR} = \max|y_m|^2 / E[|y_m|^2],\ m \in (0, N + N_g) \text{ (digital)} \tag{5.146}$$

$$\text{PAPR}_{\text{db}} = 10 \log_{10} \{\max (y(t)^2) / \text{mean}\ (y(t)^2)\},\ \text{over } 0 < t < T, \text{ (analog)} \tag{5.147}$$

T is the OFDM symbol. In general, the analog PAPR is not same as the PAPR of IFFT samples, because of the use of D/A converter. Analog PAPR is higher than digital. To bring compatibility between them, over sampling of the time domain signal is done. As an example, the base band carriers are centered at DC and the size of the IFFT can be made 8–10 times greater than the number of carriers N.

Summerizing all discussions, the following key advantages and disadvantages for OFDM can be pointed out:

Advantages

1. Efficient spectrum usage
2. Resilient to frequency selective channel
3. Effect of ISI and ICI can be completely removed by OFDM i.e robustness to channel fading and very high multi-path fading

Disadvantages

1. Perfect synchronization is required, which is difficult to achieve
2. Peak-to-average power ratio is high

Parameters for designing an OFDM System

1. Number of subcarriers
2. Guard time
3. Symbol duration
4. Subcarrier spacing
5. Modulation type per subcarrier
6. The type of forward error correction coding

Choice of parameters is influenced by system requirements such as available Bandwidth, Required bit rate, Tolerable delay spread, and Doppler values. With an example, the design of OFDM parameters can be explained as follows.

Let the initial requirements for the system is
Bit Rate ≥ 20 Mbps, Tolerable rms delay spread ≥ 200 ns, Bandwidth ≤ 15 MHz

Choose Guard time as four times the delay spread and the OFDM symbol duration as five times the guard time for 1 dB guard loss.

So, Guard Time = TG = 800 ns, TFFT = 3.2 μs, Subcarrier spacing = 1/TFFT = 312.5 kHz, 48 subcarriers in 15 MHz: 64 point IFFT/FFT for modulation/demodulation
Next choose modulation and coding rate to fulfill data rate requirement
16-QAM with rate 1/2 coding gives 24 Mbps = (4 × 1/2 × 8/4 μs)
QPSK with rate 5/6 coding gives 20 Mbps = (4 × 5/6 × 48/4 μs)

Thus, 16-QAM would be preferable because it can tolerate more weak subcarriers because of the lower coding rate.

OFDM is a very useful technique for high speed data transmission in mobile communications due to various advantages described above. Several industrial standards based on OFDM have emerged, such as Terrestrials Digital video Broadcast (DVT_T), IEEE 802.11a/g Local Area Networks (WLANs)as well as IEEE 802.16d/e broadband wireless access standards namely WiMAX.

5.23 ANALYSIS OF MODULATED SIGNALS USING VECTOR SIGNAL ANALYZER (VSA 89600) FOR ERROR VECTOR MAGNITUDE (EVM) AND RELATIVE CONSTELLATION ERROR (RCE)

There are two digital signal performance specifications: Error Vector Magnitude (EVM) and Signal to Noise Ratio (SNR). EVM (sometimes also called Received Constellation Error or RCE) is a measure used to quantify the performance of a digital radio transmitter or receiver. A signal sent by an ideal transmitter or received by a receiver would have all constellation points precisely at the ideal locations, however various imperfections in the implementation (such as carrier leakage, low image rejection ratio, phase noise, etc.) cause the actual constellation points to deviate from the ideal locations. Informally, EVM is a measure of how far the points are from the ideal locations.

In digital modulation, the base band is separated into two separate independent components, In Phase (I) and Quadrature Phase (Q). When these two components are combined they form the base band modulating signal.

In practice, Vector Modulation (I-Q modulation), the amplitude and phase will vary to generate the symbol. Each modulation scheme in vector modulation has different constellation diagram. Vector modulated symbol can be represented by vector diagram which is generated in result of modulation. The vector symbol is a graphical representation of the carrier's magnitude and phase synchronous with the symbol clock. The vector symbol is shown in Fig. 5.67.

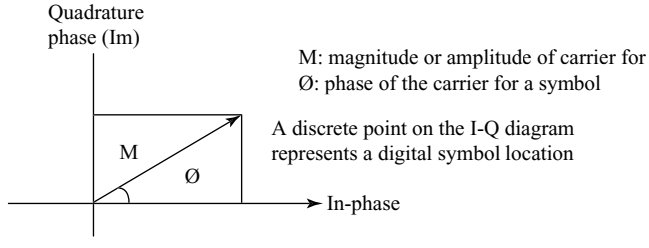


Fig. 5.67 *Vector diagram of a symbol in I-Q Modulation*

Four pieces of information are collected from the diagram as follows:

1. The distance from the *x* axis gives amplitude of in-phase component.
2. The distance from the *y* axis gives amplitude of out-phase component.
3. The length of the line joining the point from the origin gives peak amplitude of the signal.
4. The angle produced by the line with *x* axis is the phase of the signal.

After vector modulation, modulated signal (more accurately symbols) can be shown in different graphical representations like constellation diagram and eye diagram. From graphical representation, various types of statically analysis can be done to gain in-depth knowledge about the system and signal like error vector magnitude (EVM).

5.23.1 Constellation Diagram

The constellation diagram shows the graphical representation of a signal modulated by a digital modulation scheme such as Quadrature amplitude modulation or phase-shift keying. It displays the signal as a two-dimensional scatter diagram in the complex plane at all possible symbol sampling instants. In a more abstract sense, it represents the possible symbols that may be selected by a given modulation scheme as points in the complex plane. In a more compact way, we can say that constellation diagram is a graphical representation in all possible symbol location in 2-D space.

Constellation diagram also provides a graphical representation of the complex envelope of each possible symbol states. The x-axis of a constellation diagram represents the in-phase component of the complex envelope, and the y-axis represents the Quadrature component of the complex envelope. The distance between signals on a constellation diagram indicates how different the modulation waveforms are, and how well a receiver can differentiate between all possible symbols in presence of noise.

The constellation diagram for some modulation schemes are shown below.

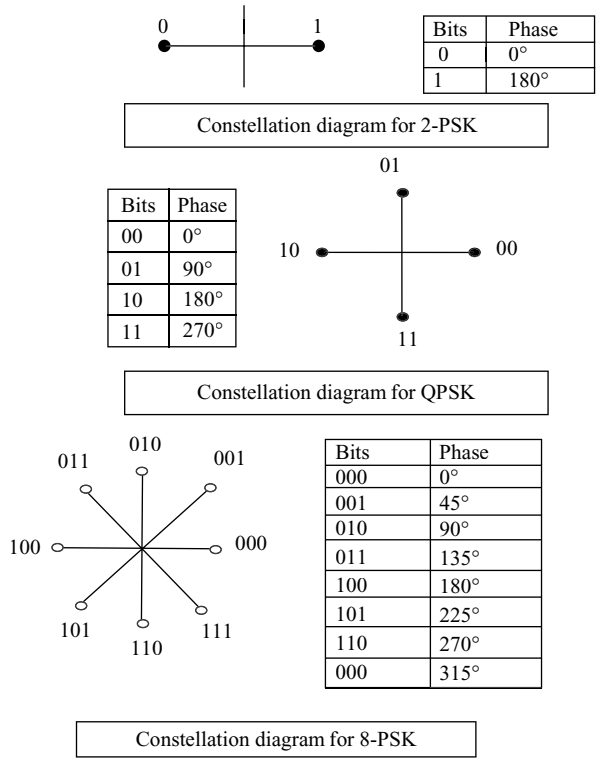


Fig. 5.68 *Constellation diagrams for different modulations*

The digital data streams are either 1 or 0. The bits can be considered as single, paired or a triplet or more depending upon the data rate required for the transmission. Measured constellation diagrams can be used to recognize the type of interference and distortion in a signal. To minimize the bit error rate during high rate data transmission, the possibility of the neighboring bits getting corrupted is more. Critical observation shows that the points in the constellation differ among themselves by a single bit. For example, the QPSK plot shows each of the constellation point (00, 01, 10 and 11) differ from its neighbor by a single bit. However, there is no restriction on position of the constellation, i.e., 10 might have a neighbor 11 or 00 but will never have 01 as its neighbor.

5.23.2 Error Vector Magnitude

The Error Vector Magnitude (EVM) is a new way of measuring the signal quality in noisy environment. EDGE (Enhanced Data Rates for GSM evolution) is the evolution of modern cellular communication networks from the Global System for Mobile (GSM). EDGE employs 8-phase shift keying (8-PSK) modulation scheme. In such a system, the Figure-of-Merit for the modulation accuracy is the EVM, which represents the distance between measured and ideal modulated signal vectors. EVM encompasses the effects caused by magnitude and phase distortions. Transmitter EVM can be measured by specialized equipment, which demodulates the received signal in a similar way to how a real radio demodulator does it.

Error vector is the difference between the measured vector with noise and the original vector. EVM, which is defined as the magnitude of the error vector normalized to the magnitude of the original vector. Thus, EVM is defined as

$$EVM_{RMS} = \sqrt{\frac{\sum_{k=0}^{N-1} |e(k)|^2}{\sum_{k=0}^{N-1} |s(k)|^2}} \tag{5.148}$$

Where N is total number of symbols, $e(k)$ is the error vector formed by the difference between the reference vector $s(k)$ and received symbol vector $r(k)$ in noisy environment. RMS-average of the phase and magnitude error experienced by multiple constellation point can be defined as RCE. Now EVM_{RMS} directly related to varying Channel SNR as

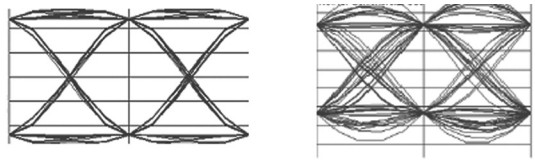
$$EVM_{RMS} \approx \left[\frac{1}{SNR}\right]^{\frac{1}{2}} = \left[\frac{N_0}{E_S}\right]^{\frac{1}{2}} \tag{5.149}$$

where E_S is the symbol energy and N_0 being the noise.

5.23.3 Eye Pattern

In telecommunication, an Eye pattern, also known as an Eye diagram, is an oscilloscope display in which a digital data signal from a receiver is repetitively sampled and applied to the vertical input, while the data rate is used to trigger the horizontal sweep. It is so called because, for several types of coding, the pattern looks like a series of eyes between a pair of rails. Eye pattern is an experimental tool to evaluate the performance of baseband pulse transmission in a noisy channel and the effect of ISI.

Several system performance measures can be derived by analyzing the display. If the signals are too long, too short, poorly synchronized with the system clock, too high, too low, too noisy, or too slow to change, or have too much undershoot or overshoot, this can be observed from the eye diagram. An open eye pattern corresponds to minimal signal distortion. Distortion of the signal waveform due to ISI and noise appears as closure of the eye pattern as shown in Fig. 5.69.



Wide opening of Eye, Low ISI Narrow opening of Eye, High ISI

Fig. 5.69 *Eye diagrams*

There are many measurements that can be obtained from an Eye diagram such as

1. Amplitude measurements for
   (a) Eye amplitude
   (b) Eye crossing amplitude
   (c) Eye crossing percentage
   (d) Eye height
   (e) Eye level
   (f) Eye SNR
   (g) Quality factor
   (h) Vertical eye opening
2. Time Measurements for
   (a) Deterministic jitter
   (b) Eye crossing time
   (c) Eye delay
   (d) Eye fall time
   (e) Eye rise time
   (f) Eye width
   (g) Horizontal eye opening
   (h) Peak–to–Peak jitter
   (i) Random jitter
   (j) RMS jitter
   (k) Total jitter
3. Interpreting measurements

Eye-diagram feature	Measurement parameter
Eye opening (height, peak to peak)	Additive noise in the signal
Eye overshoot/undershoot	Peak distortion due to interruptions in the signal path
Eye width	Timing synchronization & jitter effects

5.23.4 Performance Measurement of Modulation Schemes Using the Vector Signal Generator and Vector Signal Analyzer

Vector Signal Generator (VSG) is used to generate the vector modulated signal. Agilent's E4438C ESG vector signal generator combines outstanding RF performance and sophisticated baseband generation to deliver calibrated test signals at baseband, IF, and RF frequencies up to 6 GHz. Offering an internal baseband generator with arbitrary waveform and real-time I/Q capabilities, ample waveform playback and storage memory, and a wide RF modulation bandwidth, the E4438C ESG is equipped to test today's complex wireless systems and their components.

Vector Signal Analyzer (89600) (VSA) is a software, runs on a PC and works with a variety of hardware measurement platforms. These platforms include the 89600 VXI-based vector signal analysis systems;

the 89650S vector signal analyzer,high Performance Spectrum Analyzers (PSA), general-purpose spectrum analyzers (ESA), and the E4406A transmitter tester. These platforms down convert and digitize the signal, provide signal capture capability, and move the data to the PC in a sequential stream of data blocks. The 89600 VSA software processes the data in the time, frequency and modulation domains.

Functionalities of VSA VSA takes in amplitude vs. time data from source and then performs the following tasks.

1. Demodulates (if requested):
   (a) Analog demodulation (AM, FM, or PM)
   (b) I/Q (digital) demodulation (PSK, FSK, or QAM)
2. Performs analysis
   (a) Directly in the time domain
   (b) In the frequency domain via a Fast Fourier transform (FFT)
3. Formats the resulting data for display
   (a) Rectangular or polar coordinates
   (b) Log or linear scaling
   (c) I/Q demodulation formats
   (d) Vector and constellation diagrams
   (e) Error vector: magnitude statistics, vs. time, spectrum
   (f) I/Q errors, phase error, magnitude error
   (g) Recovered symbols

There are several analysis tools for VSA, such as

1. Vector diagram
2. Constellation diagram
3. Eye diagram
4. Demodulated spectrum
5. Error vector magnitude versus time
6. Error vector spectrum
7. Symbol table/error summary

5.23.5 Measurement Set-up with VSG and VSA for QPSK Modulation

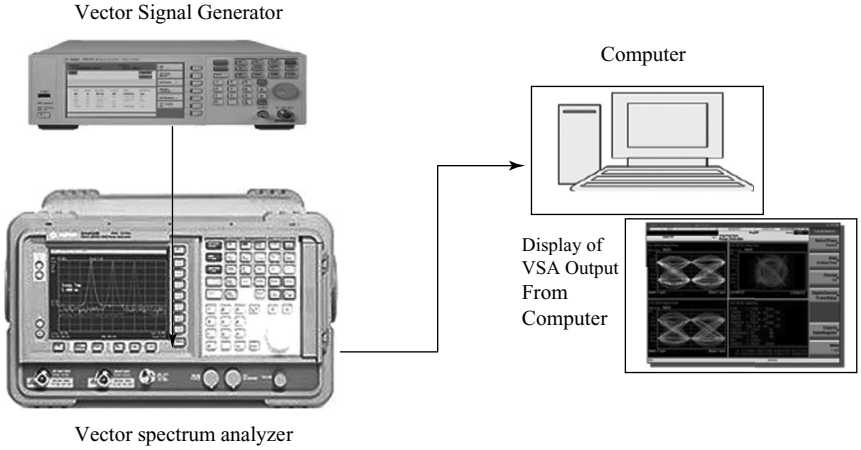


Fig. 5.70 *Interconnection of system infrastructures with VSA for measurement* *(continued)*

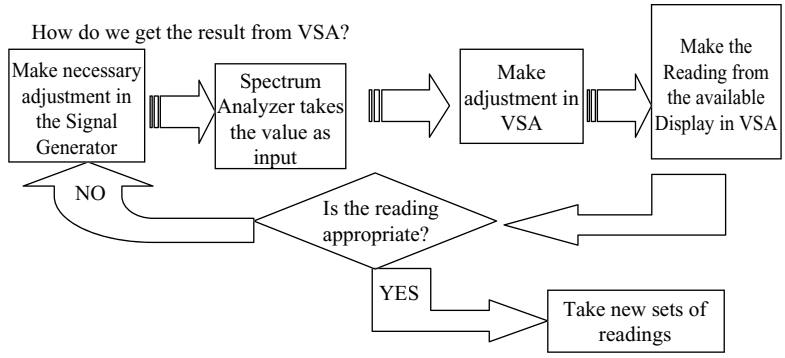
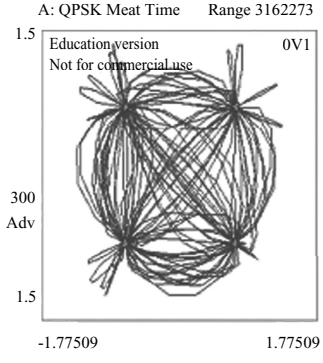


Fig. 5.70 *(continued)*

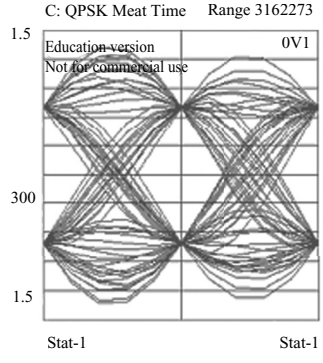
Case(a): Roll-off factor for shaping filter $\alpha = 0.2$

Experimental results for QPSK with different roll-off factors:

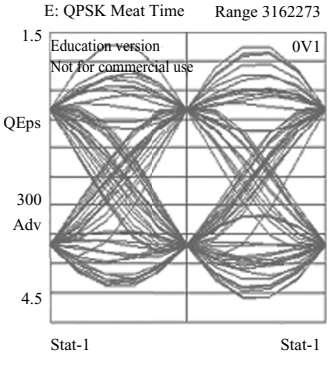
Parameter input to VSG
PN Sequence : PN–23
Filter : Root Nyquist
Symbol Rate : 25 kbps
Modulation type : QPSK
Carrier Frequency : 10 MHz
Amplitude : 0 dbm



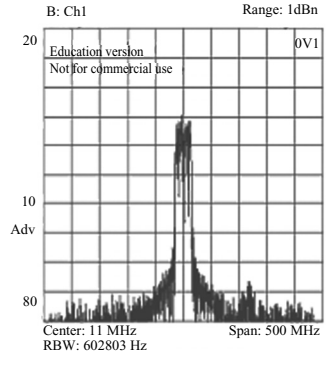
Complex received signal



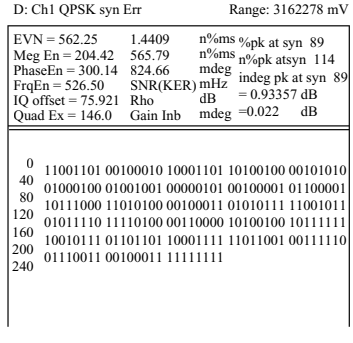
Inphase component of Eye pattern



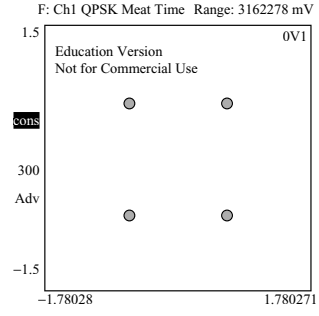
Quadrature component of Eye pattern



Occupied bandwidth of the QPSK spectrum

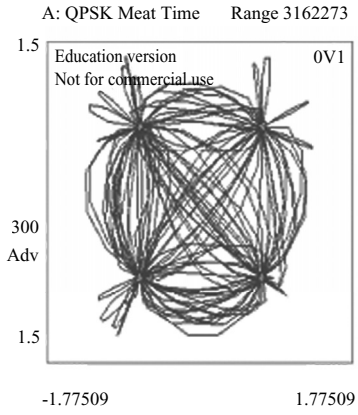


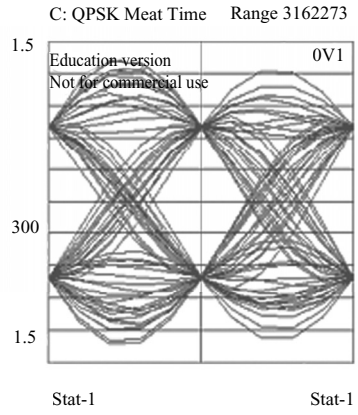
Bit patterns for demodulated signal along with error statistics EVM

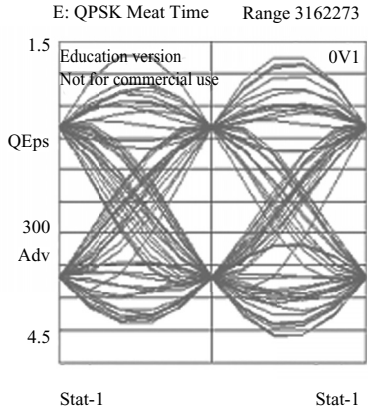


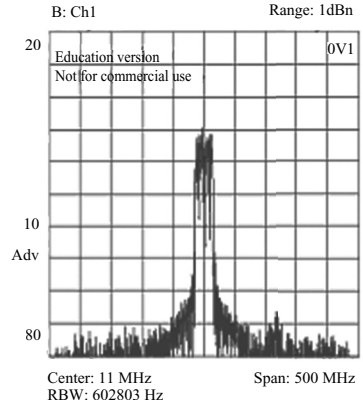
QPSK constellation diagram after filter

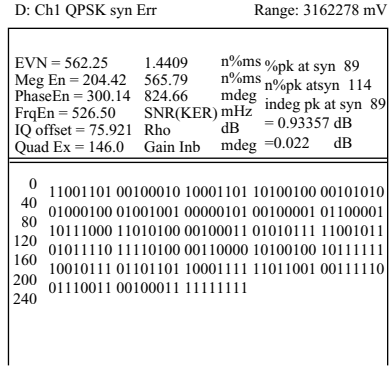
Case(b): Roll-off factor for shping filter $\alpha = 0.35$

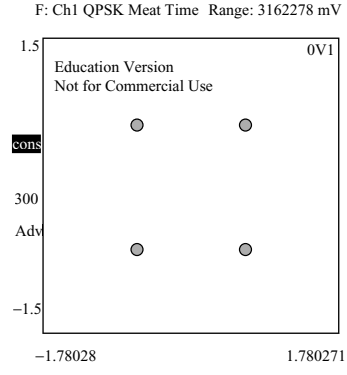












Case(c): Roll-off factor $\alpha = 0.8$

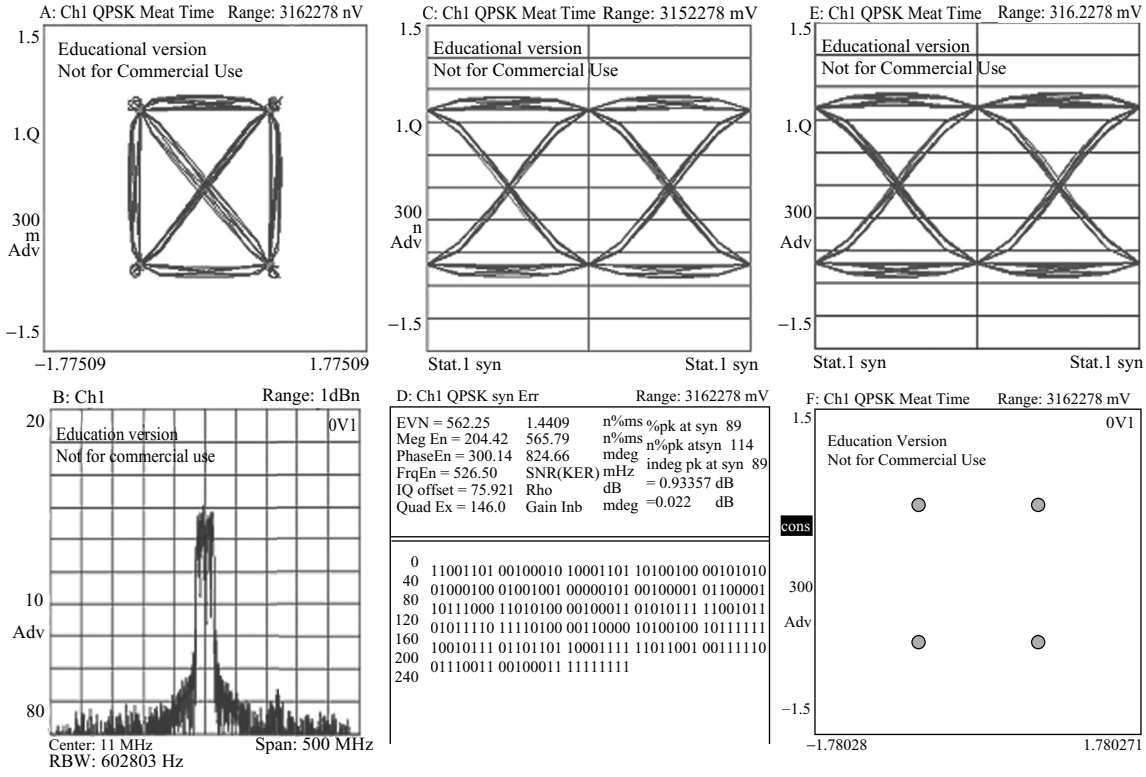


Fig. 5.72 *VSA output for different parameters of QPSK modulation with different Roll-off factor α*

Problem 5.1 The following waveforms represent the four signals $s_1(t)$, $s_2(t)$, $s_3(t)$ and $s_4(t)$. Using the Gram–Schmidt Orthogonalization procedure,

(a) find an orthonormal basis functions, for this set of signals.
(b) construct the correponding signal space diagram.

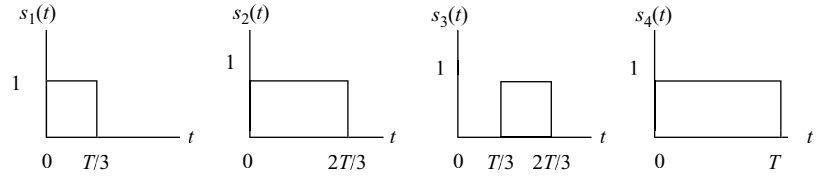


Fig. P5.1a

Solution Let us compute the energy of signal $s_1(t)$ and $s_2(t)$,

$$E_1 = \int_0^T s_1(t)^2 \, dt = \int_0^{T/3} 1^2 dt = T/3$$

$$E_2 = \int_0^T s_2(t)^2\ dt = \int_0^{2T/3} 1^2 dt = 2T/3$$

The first basis function $\varphi_1(t) = s_1(t)/\sqrt{E_1} = \begin{cases} \sqrt{3/T}, & 0 \le t \le T/3 \\ 0, & \text{Otherwise} \end{cases}$

We know that $s_{21} = \int_0^T s_2(t)\ \varphi_1(t)\ dt = \int_0^{T/3} 1(\sqrt{3/T})\ dt = \sqrt{T/3}$

So, the second basis function $\varphi_2(t) = \dfrac{s_2(t) - s_{21}(t)\ \varphi_1(t)}{\sqrt{E_2 - s_{21}{}^2}}$

$$\varphi_2(t) = \begin{cases} \sqrt{3/T}, & T/3 \le t \le 2T/3 \\ 0, & \text{Otherwise} \end{cases}$$

Again we know that $s_{31} = \int_0^T s_3(t)\ \varphi_1(t)\ dt = 0$

$$s_{32} = \int_0^T s_3(t)\ \varphi_2(t)\ dt = \int_{T/3}^{2T/3} 1.\sqrt{3/T}\, dt = \sqrt{T/3}$$

In general $g_i(t) = s_i(t) - \sum_{j=1}^{i-1} s_{ij}\ \varphi_j(t)$

So, $g_3(t) = s_3(t) - s_{31}\ \varphi_1(t) - s_{32}\ \varphi_2(t) = 1 - 0 - \sqrt{3/T}.\sqrt{T/3} = 0$ (as $\varphi_2(t)$ is 0 over $2T/3$), $\varphi_3(t)$ is indeterminant.

$$g_4(t) = s_4(t) - s_{41}\ \varphi_1(t) - s_{42}\ \varphi_2(t) - s_{43}\varphi_3(t)$$

$$g_4(t) = \begin{cases} 1, & 2T/3 \le t \le T \\ 0, & \text{Otherwise} \end{cases}$$

$$\varphi_4(t) = g_4(t)/\sqrt{\int_0^T g_4(t)^2\ dt} = \begin{cases} \sqrt{3/T}, & 2T/3 \le t \le T \\ 0, & \text{Otherwise} \end{cases}$$

Finally, with $i = 4$, the orthogonalization process is complete. So the three orthonormal basis functions are shown below.

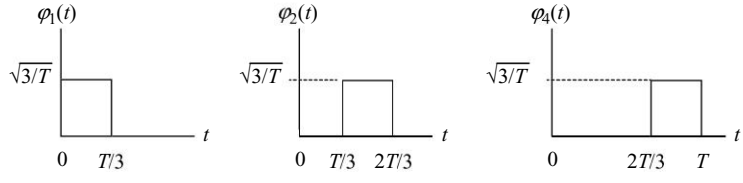


Fig. P5.1b

$M = 4$, $N = 3$, so the signals are not linearly independent.
(b) The signal space diagram is shown in Fig. P5.1c.

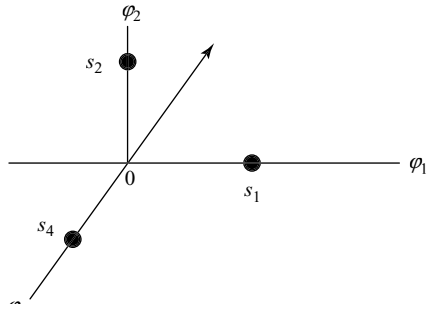


Fig. P5.1c

Problem 5.2 A received binary NRZ signal assumes the voltage levels of 500 mV and –500 mV for bit transmission of '1' and '0' respectively. The bit rate is r bits/sec. The signal is corrupted by additive white Gaussian noise with a two-sided spectral density of 10^{-6} volts^2/Hz. The received signal is processed by an Integrate and dumps circuit in every bit interval and compared with zero thresholds to take a bit decision.

Assuming '1' and '0' transmission to be equally likely, the maximum value of 'r' is such that the bit error probability $\leq 10^{-5}$

Given $1/\sqrt{2\pi}\int_{x}^{\infty} e^{-z^2/2}\,dz = 10^{-5}$ at $x = 4.27$ and erfc $(x) = 2/\sqrt{\pi}$

and

$$\text{erfc}(x) = 2/\sqrt{\pi}\int_{x}^{\infty} e^{-z^2/2}\,dz = 2\times10^{-5} \text{ at } x = 3.02$$

Find the bit duration, the energy per bit and BER in terms of E_b/N_0 ratio.

Solution

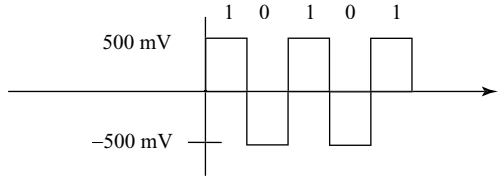


Fig. P5.2

The transmitted signal $s_1(t) = 500$ mV and $s_2(t) = -500$ mV over $0 \leq t \leq T$

The energy difference of the signal $E_g = \int_0^T [s_1(t) - s_2(t)]^2\,dt = (2A)^2 T = 4A^2T_b$

Bit error probability $= P_e = Q(\sqrt{E_g/2N_0}) = Q\left(4A^2T_b/2N_0\right) \leq 10^{-5}$

$$4A^2T_b/2N_0 = 4.27, N_0/2 = 10^{-6}\,\text{volt}^2/\text{Hz} \Rightarrow N_0 = 2\times10^{-6}$$

$$\Rightarrow T_b = 4.27\times2N_0/4A^2 = 4.27\times4\times10^{-6}/4\times(500\times10^{-3})^2 = 4.27/250000$$

$$= 1.708\times10^{-5} \text{ sec}$$

As given in Eqs. (5.36) to (5.42), for this special case of $p_{01} = p_{10}$, the channel is binary symmetric and the average probability of symbol error p_e becomes,

$$p_e = 1/2 \text{ erfc}\left(A/\sqrt{N_0/T_b}\right)$$

The transmitted symbol energy per bit is

$$E_b = A^2 T_b = \left(500 \times 10^{-3}\right)^2 x\ 1.708 \times 10^{-5}$$
$$= 250000 \times 1.708 \times 10^{-11} = 42.7 \times 10^{-7}$$

We finally formulate the p_e in terms of transmitted symbol energy per bit to the noise spectral density E_b/N_0 ratio solely.

$$1/2 \text{ erfc}\sqrt{(4.27\text{x}10^{-7} / 2\text{x}10^{-6})} P_e = 1/2 \text{ erfc}\left(\sqrt{E_b/N_0}\right) = 1/2 \text{ erfc}\sqrt{(A^2 T_b / N_0)}$$
$$= 1/2 \text{ erfc}\sqrt{(4.27/2)}$$
$$= 1/2 \text{ erfc}\ (1.46) = 0.9598/2 = 0.4799$$

Problem 5.3 An orthogonal set of signals is characterized by the property that the inner product of any pair of signals in the set is zero. The following Figure is the set of signals $s_1(t)$ and $s_2(t)$ satisfy the orthogonality condition. Construct the signal constellation for $s_1(t)$ and $s_2(t)$.

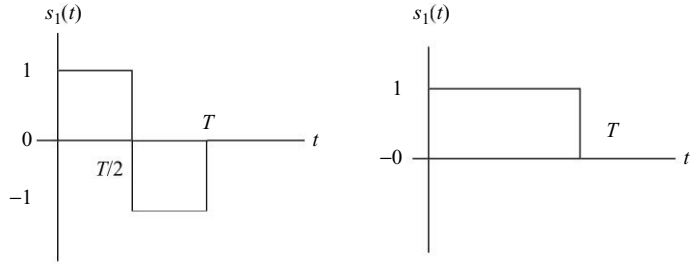


Fig. P5.3a

Solution Let us compute the energy of the signals.

$$E_1 = \int_0^T S_1(t)^2\, dt = \int_0^{T/2} (1)^2\, dt + \int_{T/2}^{T} (-1)^2 T/2 + T/2 = T$$

$$E_2 = \int_0^T S_2(t)^2\, dt = \int_0^T 1^2\, dt = T$$

To represent the orthogonal signals $s_1(t)$ and $s_2(t)$, we need two basis functions. The first basis function is $\varphi_1(t) = s_1(t)/\sqrt{E_1} = s_1(t)/\sqrt{T}$

So, $s_1(t) = \sqrt{T}\ \varphi_1(t), \quad 0 \le t \le T/2$

Given that $s_1(t)$ and $s_2(t)$ are orthogonal functions,

We know that $s_{21} = \int_0^T s_2(t)\, \varphi_1(t)\, dt = \int_0^{T/2} 1\,(\sqrt{1/T})\, dt + \int_{T/2}^{T} (-1)\,(\sqrt{1/T})\, dt = 0$

A new intermediate function can be introduced as

$$g_2(t) = s_2(t) - s_{21}(t)\varphi_1(t),$$

Second basis function $\varphi_2(t) = g_2(t) \Big/ \sqrt{\int_0^T g^2{}_2(t)\, dt}$

$$= \frac{s_2(t) - s_{21}(t)\,\varphi_1(t)}{\sqrt{E_2 - s_{21}{}^2}} = s_2(t)/\sqrt{E_2} = 1/\sqrt{T}$$

$$\varphi_2(t) = \begin{cases} 1/\sqrt{T} & 0 \le t \le T \\ 0, & \text{Otherwise} \end{cases}$$

So, $s_2(t) = \sqrt{T}\ \varphi_2(t)$

The signal constellation diagram for $s_1(t)$ and $s_2(t)$ is

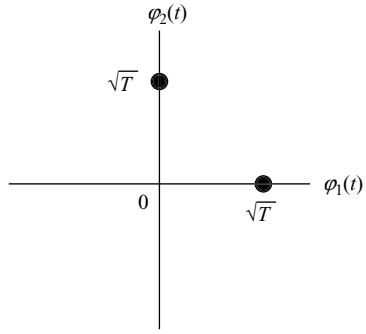


Fig. P5.3b

Problem 5.4 An 8-level PAM signal is defined by

$s_i(t) = A_i$ rect $(t/T - t/2)$, where $A_i = \pm 1, \pm 2, \ldots . \pm 7$. Formulate the signal constellation for $\{s_i(t)\}^8{}_{i=1}$

Solution The energy of the signals $s_i(t)$ is given by

$$E_i = \int_0^T S_i(t)^2 dt = \int_0^T A_i{}^2 dt = A_i{}^2 T, A_i = \pm 1, \pm 2, \ldots . \pm 7$$

The basis function is given by $\varphi_1(t) = s_i(t)/\sqrt{E_i} = s_i(t)/A_i\sqrt{T}$

$$s_i(t) = A_i\ \sqrt{T}\ \varphi_1(t)$$

Therefore, the signal space diagram of the 8-level PAM signal is as follows:

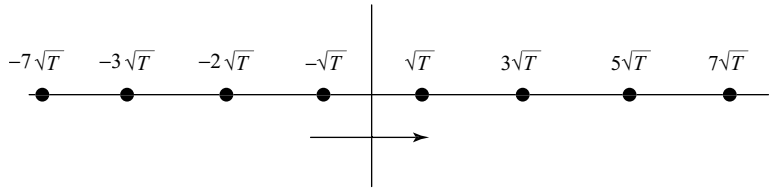


Fig. P5.4

Problem 5.5 Using Gram–Schmidt orthogonalization procedure, find a set of orthonormal basis functions to represent the three signals $s_1(t)$, $s_2(t)$, and $s_3(t)$ shown in the Figure. Express each of these signals in terms of the set of basis functions obtained.

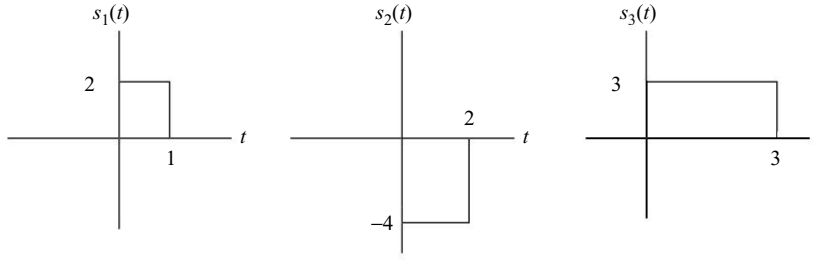


Fig. P5.5a

Solution The energy of $s_1(t)$ is

$$E_1 = \int_0^1 S_1(t)^2\,dt = \int_0^1 2^2\,dt = 4$$

The first basis function,

$$\varphi_1(t) = s_1(t)/\sqrt{E_1} = 2/\sqrt{4} = \begin{cases} 1, & 0 \le t \le 1 \\ 0, & \text{Otherwise} \end{cases}$$

$$s_{21} = \int_0^1 s_{21}(t)\,\varphi_1(t)\,dt = \int_0^1 -4.1\,dt = -4$$

$$g_2(t) = s_2(t) - s_{21}(t)\,\varphi_1(t) = \begin{cases} -4, & 1 \le t \le 2 \\ 0, & \text{Otherwise} \end{cases}$$

Second basis function $\varphi_2(t) = g_2(t)/\sqrt{\int_0^T g^2{}_2(t)\,dt} = -4/\sqrt{\int_1^2 (-4)^2\,dt} = \begin{matrix} -1, & 1 \le t \le 2 \\ 0, & \text{Otherwise} \end{matrix}$

Now,

$$s_{31} = \int_1^2 s_3(t)\,\varphi_1(t)\,dt = 3$$

$$s_{31} = \int_1^2 s_3(t)\,\varphi_2(t)\,dt = -3$$

$$g_3(t) = s_3(t) - s_{31}\,\varphi_1(t) - s_{32}\,\varphi_2(t) = \begin{matrix} 3, & 0 \le t \le 3 \\ 0, & \text{Otherwise} \end{matrix}$$

Hence, the third basis function is

$$\varphi_3(t) = g_3(t)/\sqrt{\int_0^T g^2{}_2(t)\,dt} = 3/\sqrt{\int_0^2 (3)^2\,dt} = \begin{matrix} 1, & 2 \le t \le 3 \\ 0, & \text{Otherwise} \end{matrix}$$

The three basis functions are represented as follows:

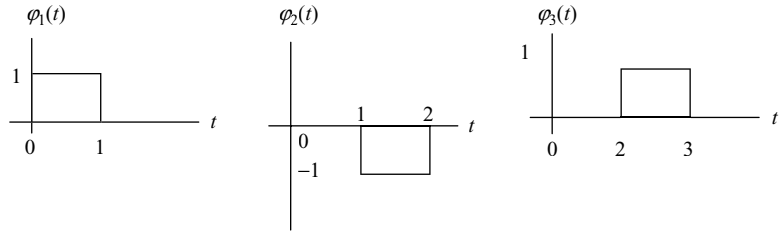


Fig. P5.5b

(b) $s_1(t) = s_{11}\,\varphi_1(t) = 2\,\varphi_1(t)$
$s_2(t) = s_{21}\,\varphi_1(t) + s_{22}\,\varphi_2(t) = -4\,\varphi_1(t) + 4\,\varphi_2(t)$
$s_3(t) = s_{31}\,\varphi_1(t) + s_{32}\,\varphi_2(t) + s_{33}\,\varphi_3(t) = 3\varphi_1(t) - 3\varphi_2(t) + 3\varphi_3(t)$

Problem 5.6 A pair of signals $s_i(t)$ and $s_k(t)$ have a common duration T. show that the inner product of this pair of signals is given by

$$\int_0^T s_i(t)\, s_k(t)\, dt = s_i^{Ts}{}_k$$

where s_i and s_k are the vector representation of $s_i(t)$ and $s_k(t)$, respectively.

(b) Also show that, $\int_0^T (cs_i(t) - s_k(t))^2\, dt = \| s_i - s_k \|^2$

Solution The pair of signals belonging to an N-dimensional signal space can be represented as linear combination of N orthogonal basis functions. Thus it follows:

$$s_i(t) = \sum_{j=1}^{N} s_{ij}\, \varphi_j(t), \qquad 0 \le t \le T$$

where the coefficients of the expansion are defined by

$$s_{ij} = \int_0^T s_i(t)\, \varphi_j(t)\, dt, \;\; i = 1,2 \text{ and } j = 1,2$$

The real valued basis functions $\varphi_i(t)$ and $\varphi_j(t)$ are orthogonal.
Hence

$$\int_0^T \varphi_i(t)\, \varphi_j(t)\, dt = \partial_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0, & \text{Otherwise} \end{cases}$$

Now the set of coefficients $\{s_{ij}\}_{j=1}^{N}$ may be viewed as an N-dimensional vector defined by

$$s_i = \begin{bmatrix} s_{i1} \\ s_{i2} \\ \vdots \\ s_{iN} \end{bmatrix}, i = 1,2,\ldots M$$

where M is the number of signals in the set with $M \ge N$.
Therefore, the inner product of the pair of signals $s_i(t)$ and $s_k(t)$ is given by

$$\int_0^T s_i(t)\, s_k(t)\, dt$$

$$= \int_0^T \left[\sum_{j=1}^{N} s_{ij}\ \varphi_j(t)\right]\left[\sum_{1=1}^{N} s_{kl}\ \varphi_1(t)\right] dt$$

$$= \sum_{j=1}^{N} s_{ij} \sum_{1=1}^{N} s_{kl} \int_0^T \varphi_j(t)\, \varphi_1(t)\, dt$$

$$= \sum_{j=1}^{N} s_{ij}\ s_{kj} = s_i^T s_k$$

(b)

$$\int_0^T (s_i(t) - s_k(t))^2\, dt$$

$$= \int_0^T \left\{(s_i^2(t) + s_k^2(t))^2 - 2\, s_i(t) s_k(t)\right\} dt$$

$$= \int_0^T \left\{(s_i^2(t) + s_k^2(t))^2 - s_i(t) s_k(t) - s_k(t) s_i(t)\ \right\} dt$$

$$= s_i^T s_i + s_k^T s_k - s_i^T s_k - s_k^T s_i$$

$$= (\ s_i - s_k)^T (s_i - s_k) = \|\ s_i - s_k\ \|^2$$

Problem 5.7 In the Baye's test, applied to a binary hypothesis-testing problem where we have to chose one or two possible hypothesis H_0 and H_1, we minimize the risk R defined by

$$\begin{aligned} R = {} & C_{00}\, p_0\, P(\text{say } H_0 \mid H_0 \text{ is true}) \\ & + C_{10}\, p_0\, P(\text{say } H_1 \mid H_0 \text{ is true}) \\ & + C_{11}\, p_1\, P(\text{say } H_1 \mid H_1 \text{ is true}) \\ & + \mathrm{C}_{01}\, p_1\, P(\text{say } H_0 \mid H_1 \text{ is true}) \end{aligned}$$

The terms C_{00}, C_{10}, $C_{11,}$ and C_{01} denote the costs assigned to the four possible outcomes of the experiment: The first subscript indicates the hypothesis chosen, and the second the hypothesis is true. Assume that $C_{10}>C_{00}$ and $C_{01}>C_{11.}$ The p_0 and p_1 denote the *a priori* probabilities of hypothesis H_0 and H_1, respectively.

Solution

(a) Given the observation vector x, show that the partitioning of the observation space so as to minimize the risk R leads to the likelihood ratio test,

$$\begin{aligned} &\text{Say } H_0 \text{ if } \Lambda(x) < \lambda \\ &\text{Say } H_1 \text{ if } \Lambda(x) > \lambda \end{aligned}$$

where $\Lambda(x)$ is the likelihood ratio,

$$\Lambda(\mathrm{x}) = f_x(x \mid H_1) / f_x(x \mid H_0)$$

and λ is the threshold of the test defined by, $\lambda = p_0\ (C_{10} - C_{00}) / p_1\ (C_{01} - C_{11})$

(b) What are the cost values for which the Bayes' criterion reduces to the minimum probability of error criterion?

Solution Let z denotes the observation space and $z = z_0, z_1$.
Let x be the observation.
So, if $x \in z_0$ then H_0 is true and
If $x \in z_1$ then H_1 is true.
The probability $P_{Ho} = f_{x|Ho}$ and $P_{H1} = f_{x|H1}$
So the risk function,

$$R = C_{00} p_0 \int z_0 \, fx \,|_{Ho}\,(x / H_0)dx + C_{10}\, p_0 \int z_1 f_x \,|\, H_0(x / H_0)dx$$
$$+ \; C_{11} p_1 \int z_1 f_x \,|_{H1}\,(x / H_1)dx + C_{01}\, p_1 \int z_0 \, f_x \,|\, _{H1}(x / H_1)dx$$

For Baye's test, R will be minimum. Keeping this constraint, we have to construct the decision rule for x to be either in z_0 or z_1.
Rewriting the risk factor, considering $z = z_0 + z_1$

$$R = C_{00} P_0 \int Z_0 f_x \left| H_0(x / H_0)dx + C_{10} p_0 \int z_1 f_x \,|h_0(x / H_0)\, dx \right|$$
$$+ C_{11} p_1 \int (z - z_0) f_x \left| H_1(x / H_1)dx + C_{01} p_1 \int Z_0 f_x \right|_{H1}(x / H_1)\, dx$$

Again z is the universal space, so P_z is a sure event =1.

So, $\int z f_x \,|\, _{Ho}(x / H_0)dx = z\, f_x \,|\, _{H1}(x / H_1)dx \; = 1$
Risk factor R reduces to

$$R = (p_0 C_{10}) + p_1 C_{11}) + \int z_0 \{-[p_0 (C_{10} - C_{00})\, f_x \,|\, _{Ho}(x / H_0)]$$
$$+ \; [p_1 (C_{01} - C_{11})\, f_x \,|\, _{\mathrm{H1}}\,(x | H_1)]\}\, dx$$
$$= \textbf{Fixed cost + I + II}$$

$C_{10} > C_{00}$ and $C_{01} > C_{11}$ (given), so I and II are positive.
But if I>II then $x \in z_0$ as it minimize the overall cost.
And if I < II then $x \in z_1$ as it increases the cost. Thus the decision criterion could be as follows.

If, $p_1 (C_{01} - C_{11})_{f_x|H1}(x \,|\, H_1) > p_0 (C_{10} - C_{00})\; _{f_x|H0}(x \,|\, H_0)$ then
$x \in z_1$ and H_1 is true. Otherwise $x \in z_0$ and H_0 is true

Thus, $\dfrac{f_x(x|H1)}{(x|H1)} \underset{H_0}{\overset{H_1}{\gtrless}} \dfrac{p_0 (C_{10} - C_{00})}{p_1 (C_{01} - C_{11})}$

Now, λ is defined by, $\dfrac{p_0 (C_{10} - C_{00})}{p_1 (C_{01} - C_{11})}$

i.e., the Likelihood ratio $\Lambda(x) = f_x(x|H_1) / f_x(x|H0) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$

(b) For minimum probability of error criterion, the likelihood ratio test is

$$\Lambda(x) \underset{H_0}{\overset{H_1}{\gtrless}} P_0 / P_1$$

This is hold if and only if , $C_{00} = C_{11} = 0$ and $C_{10} = C_{01}$, i.e., cost of correct decision = 0 and the cost of error of 1 type = cost of the error of other type.

Summary

Digital modulation techniques are the foundation of modern wireless communication. With the increasing demand of higher and higher data rate, the necessity of sophisticated physical layer communication techniques are also needed. Here lies the importance of understanding the basic methodologies of digital communication. In this chapter, details of the digital communications starting from signal-space analysis to different modulation schemes have been thoroughly discussed. For the useful insight of the subject, some very relevant problems are also worked out.

References

[1] Haykin, Simon, *Communication Systems*, 4th edition New York, NY. John Wiley & Sons, 2001.
[2] *What is GMSK Modulation—Gaussian Minimum Shift Keying*, www.Radio-Electronics Com.htm
[3] *What is MSK Modulation—Minimum Shift Keying*, www.Radio-Electronics Com.htm
[4] *GMSK: Practical GMSK Data Transmission,* http://www.eetchina.com/ARTICLES/2003AUG/PDF/2003AUG29_NTEK_AN.PDF
[5] *Minimum Shift Keying: A Spectrally Effiecient Modulation,* http://www.elet.polimi.it/upload/levantin/SistemiIntegrati/msk_pasupathy_1979.pdf
[6] Fazel, K., and S. Kaiser (2008), *Multi-Carrier and Spread Spectrum Systems: From OFDM and MC-CDMA to LTE and WiMAX*, 2nd Edition, John Wiley & Sons, 2008, ISBN 978-0-470-99821-2
[7] http://kom.aau.dk/~imr/RadioCommII/, Aalborg University
[8] Chandra, Athaudage , *OFDM for Wireless Multimedia Communications: Synchronization and Other Issues*, ARC Special Research Centre for Ultra-Broadband Information Networks, CUBIN.
[9] *Orthogonal frequency division multiplexing* http://www.ert.rwth-aachen.de/Projekte/Theo/OFDM/node2.html
[10] Molisch Andreas F., *Wireless Communications*, Wiley and Sons, 2005.
[11] Chang, R.W., and R.A Gibbey, *A Theoretical Study of Performance of an Orthogonal Multiplexing Data Transmission Scheme*, IEEE Transactions on Communications Technology **16** (4), pp. 529–540, 1968 .

Questions for Self-Test

5.1 A message source emits one symbol every T secs, the M symbols denoted by m_1, m_2, ….m_N and the M symbols are equally likely, what would be the probability that symbol m_i would be emitted by source?

5.2 With the message m_i, the transmitter convert it into distinct signal $s_i(t)$ over full duration T suitable for transmission over channel. What would be the energy of the signal $s_i(t)$?

5.3 During the transmission over noisy channel, signal $s_i(t)$ gets corrupted by AWGN and received $x(t)$ by the receiver. Write the expression for $x(t)$. Now in extracting the symbol m_i from the received signal $x(t)$, receiver makes erroneous decision because of the noise. What would be the average probability

of symbol error? What criterion does the receiver should satisfy in designing?

5.4 Define orthonormal set of basis functions.

5.5 In geometric representation of energy signals, what will happen if the basis functions are not orthonormal?

5.6 Relate Euclidean distance d_{ik} between the signal vectors s_i and s_k.

5.7 Express Schwarz inequality for two enrgy signals. At what condition does equality hold?

5.8 Gram-Schmidt orthogonalization procedure is the mathematical representation of the geometric representation of energy signals–explain.

5.9 What are the two main distinction of Gram-Schmidt orthogonalization with Fourier series expansion?

5.10 Consider a digital communication system that involves two equally likely messages represented by two signal vectors s_i and s_k. Find the probability of symbol error averaged over for all two symbols in terms of Euclidean distance d_{ik}.

5.11 What are the three basic ways of base band pulse transmission? Explain with diagrams.

5.12 Discuss about the hierarchy of digital communication systems.

5.13 What are the different types of coherent phase shift keying?

5.14 Express the signals for coherent BPSK system. Show that signals space for BPSK is one-dimensional having two message points.

5.15 In BPSK system, find the conditional probability of the receiver deciding in favor of symbol 1, given that symbol 0 was transmitted. Find the Average probability of error for this system.

5.16 Describe the process of generating and detecting BPSK signals.

5.17 Find the power spectral density of the BPSK signals. What is the bandwidth required for this system?

5.18 Find the four message points and the associated signals for QPSK. How many basis functions are required for this system.

5.19 The input binary data stream to a QPSK system is 01101000. Show the waveforms representing the two componets of QPSK signals $s_{i1}\ \varphi_1(t)$ and $s_{i2}\ \varphi_2(t)$. Then generate the QPSK signal by adding the two componets.

5.20 How is QPSK signal generated and detected?

5.21 Finding the power spectral density for QPSK signal, compare it with that of BPSK.

5.22 Explain the importance of OQPSK scheme.

5.23 Show that M-ary PSK signals are bandwidth efficient.

5.24 Draw the signal space diagram for 16-QAM signals. Is there any relationship with 4-PAM signals?

5.25 What is the probability of symbol error for M-ary QAM signals?

5.26 Minimum Shift Keying is based on the principle of continuous phase shift keying (CPSK)–explain.

5.27 Explain the name for MSK as Minimum Sfift Keying.

5.28 Show the MSK pulse shaping for 1 bit and 0 bit.

5.29 Find the in phase and quadrature components of MSK signal.

5.30 Explain the four possible forms of MSK signal in terms of $\theta(0)$ and $\theta(T_b)$. Provide pictorial representations.

5.31 How is signal space diagran formed for MSK signal? Show the I-Q components of MSK signal when 0100100-bit sequence is transmitted. Show the final MSK signal as the sum of these two components.

5.32 Find the Energy and Power spectral density for MSK signal.

5.33 Give the advantages and disadvantages of MSK over QPSK.

5.34 GMSK modulation is based on MSK, but more spectral efficient than MSK–explain.

5.35 OFDM modulation is a technique for high data rate transmission–how is it possible?
5.36 OFDM is orthogonal FDM, but saves 50 % bandwidth over FDM–explain how?
5.37 Discuss about the modulation and demodulation schemes for OFDM in analog technique.
5.38 What is called Cyclic Prefix? What is the relationship of CP with guard interval?
5.39 How is ICI and ISI eliminated by the use of CP?
5.40 Define Error Vector Magnitude. How does it help in quantifying signal quality?
5.41 Explain the formation of constellation diagram for 4 PSK signal.
5.42 What is the significance of Eye Diagram in communication system? How does the roll-off-factor influence the Eye Diagram of a signal system?

Equalization, Diversity and Coding for Fading Channels: The Receiver Techniques

6

Introduction

Wireless communication channel is always unreliable due to the hostile and dynamic environments through which signal propagates. Signal faces multi-path propagation, Doppler frequency shifts and delay spread causing signal degradation (fading) and the distinguishability in its original form is greatly affected resulting in higher bit error rate. Very often the system requires signal-processing techniques to improve the radio link performance.

Equalization, diversity, and channel coding techniques are three independent techniques those are used separately or in combination to process the signal to improve received signal quality and link performance within small scale time and distances.

Three different effects are the main reasons for multi-path propagation. These are reflection, diffraction and scattering. The way multi-path signals recombine at the receiver end produces the type of distortion in the received signal. When the signals with identical delays are recombined, the produced distortion known narrowband fading. But recombination with different delays causes wideband fading along with pulse spreading that leads to inter-symbol interference (ISI). **Equalization technique** compensates for ISI created by multi-path within time dispersive channels. An equalizer within a receiver compensates for the average range of expected channel amplitude and delay characteristics. Equalizers **must be adaptive** since the channel is generally unknown and time varying.

Diversity techniques can be used to improve system performance in fading channels. Instead of transmitting and receiving the desired signal through one channel, L copies of the desired signal through M different channels are received. The idea is that while some copies may undergo deep fades, others may not. We might still be able to obtain enough energy to make the correct decision on the transmitted symbol. There are several different kinds of diversity, which are commonly employed in wireless communication systems. When the equalizer is used to counter the effects of time dispersion (ISI), diversity is used to reduce the depth and duration of the fades experienced by a receiver in a local area. This may be applied both at base station and mobile station. There are several types of diversity techniques like space diversity, frequency diversity, time diversity, and polarization diversity of which space diversity is the very common to use.

Channel coding techniques improve the small-scale link performance by adding redundant bits in the transmitted message and is used to detect or correct the error occurs in any received bit. The added redundant bits reduce the overall data transmission rate. The three main coding are block coding, convolutional coding and turbo coding.

In this chapter, very basics of the principle of equalization, algorithms for adaptive equalization, diversity techniques, designing of RAKE receiver and the coding techniques are provided.

6.1 INTER SYMBOL INTERFERENCE (ISI)

Consider pulsed information is transmitted over an analog channel, such as a phone line or airwaves with impulse response $h(t)$. Even though the original signal is a discrete time sequence (or a reasonable approximation), the received signal is a continuous time signal.

The received signal output is the convolution integral of the input signal and channel response:

$$r(t) = \int_{-\infty}^{\infty} x(\tau)\, h(t-\tau)\, d\tau \;=\; \int_{-\infty}^{\infty} x(t-\tau)\; h(\tau)\, d\tau \tag{6.1}$$

where $r(t)$ is the received signal, $h(t)$ is the channel impulse response, and $x(t)$ is the input signal. The second half of the equation above is a result of the fact that convolution is a commutative operation.

Now consider a wireless communication system where at the transmitter, the modulator maps the information sequence into data sequence of $\{x_n\}$ in the N-dimensional signal space depending on the specific modulation scheme. We can model the transmitted signal $x(t)$ by the input train of pulse, which consists of periodically transmitted impulses of varying amplitudes.

$$x(t) = \sum_{-\infty}^{\infty} x_n\; \delta(\, t - nT), \text{ where } T \text{ is the symbol duration.}$$

Therefore,

$$x(t) = \; 0 \text{ for } t \neq nT$$
$$x(t) \; = x_n \text{ for } t = nT$$

This means that the only significant values of the variable of integration in the integral Eq. (6.1) are those for which $\tau = nT$, any other values of τ result the integral equal to 0. Now if the effective communication channel is viewed as the combination of transmitter filter, channel and the receiver filter in tandem as shown in Fig. 6.1 with impulse response $h_{\text{eff}}(t)$, then the output received signal of the effective channel is the convolution of the input $x(t)$ and $h_{\text{eff}}(t)$. Therefore $r(t)$ at the end of the n^{th} sample $t = nT$ can be written as

$$r(t)\,|_{t=nT} = \sum_{n=-\infty}^{\infty} x_n h_{\text{eff}}\,(t-nT) + n(nT) \tag{6.2}$$

where $n(t)$ is the AWGN introduced by the channel. In discrete time sequence with T as the time unit, $r(t)$ is represented as $r_n = r(nT)$, $h_{\text{neff}} = h_{\text{eff}}(nT)$ and $n_{\text{n}} = n(nT)$.

It is seen that the received signal consists of the sum of many scaled and shifted continuous time system impulse responses. The impulse response is scaled by the amplitudes of the transmitted pulses of $x\,(t)$.

$$r_n = r(nT) = \sum_{-\infty}^{\infty} x_n h_{\text{eff}}\,(nT-kT) + n_n = x_n\; h_{\text{eff}}\,(0) + \sum_{k=-\infty, K\neq n}^{\text{k}=\infty} \text{x}_{\text{k}}\, \text{h}_{\text{eff}}\,(nT-kT) + n_n \tag{6.3}$$

To account for an arbitrary phase offset if the sample clock is not synchronized with transmitted data then an offset t_0 *is to be added* to the time index in the above equation. Unless the sample clock is perfectly synchronized with the transmit clock, the sample-phase offset will be nonzero.

$$r(nT + t_0) = x_n h_{eff}\,(t_0) + \sum_{k\neq\infty} x_{\text{k}} h_{\text{eff}}\,(t_0 + nT - kT) + n_n \tag{6.4}$$

In the equation above, the first term is the desired signal component modified by channel gain that contains transmitted information of the n^{th} transmitted symbol and is obtained by multiplying it with the center tap of the channel-impulse response. The second product terms in the summation represents the Inter Symbol Interference (ISI) terms due to other transmitted symbols.

To minimize the probability of bit transmission error, the optimum receiver consists of matched filter, an equalizer and a maximum likelihood decision device. The matched filter is matched to the transmitter filter in tandem with physical channel, where as the equalizer is the transversal filter that compensates for ISI as shown in Fig. 6.1.

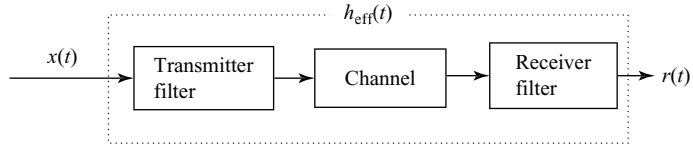


Fig. 6.1 *Equivalent channel with impulse response* $h_{\text{eff}}(t)$

6.2 EQUALIZATION TECHNIQUE

The operational principle of an equalizer can be visualized either in the time domain or in the frequency domain. The delay dispersion corresponds to frequency selectivity. The transfer function is not constant over the considered system bandwidth, as a result ISI occurs. In wireless communication, ISI is the biggest obstacle for high-speed data transmission and equalization is used to combat the ISI. The purpose of equalizer is to reverse the channel distortion. This is basically a signal processing technique to minimize ISI. The effect of an equalization system is to compensate for transmission-channel impairments such as frequency-dependent phase and amplitude distortion. Besides correcting for channel frequency-response anomalies, the equalizer can cancel the effects of multipath signal components, which can manifest themselves in the form of voice echoes, video ghosts or Rayleigh fading conditions in mobile communications channels.

Equalizer is usually implemented at baseband or at IF in a receiver as given in Fig. 6.2. It represents the communication system at the receiver site using adaptive equalization process. Let $s(t)$ is the original signal information and $f(t)$ is the combined signal for transmitter, channel and RF and IF section of the receiver site. So, the received signal at the equalizer input along with the noise part $n_g(t)$ is

$$r_{\text{in\_eq}}(t) = s(t)^* f^*(t) + n_g(t) \tag{6.5}$$

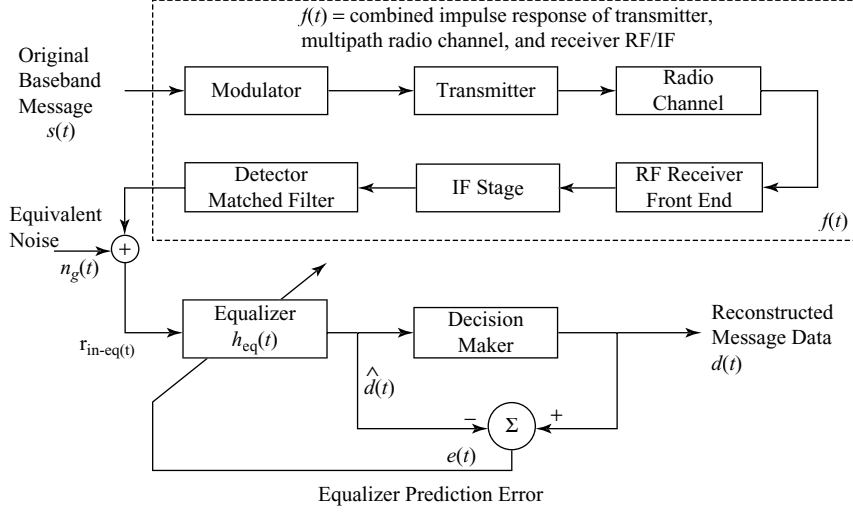


Fig. 6.2 *Block diagram of a communication system using adaptive equalizer*

where $f^*(t)$: complex conjugate of $f(t)$; $n_g(t)$: baseband noise at the input of the equalizer. If $h_{eqz}(t)$ be the impulse response of the equalizer, then the output $d(t)$ of the equalizer is the convolution of $r_{in-eq}(t)$ and $h_{eqz}(t)$.

$$d(t) = r_{in\_eq}(t)^* h_{eqz}(t) = s(t)^* f^*(t)^* h_{eqz}(t) + n_g(t)^* h_{eqz}(t) \tag{6.6}$$

Let the complex impulse response of the equalizer is $h_{eqz}(t) = \Sigma_n c_n \delta(t - nT)$, c_n is the complex filter coefficients of the equalizer. In the absence of noise, the equalizer output should be equal to the input information, i.e., $d(t) = s(t)$, in that case the desirable condition is $f^*(t) * h_{eqz}(t) = \delta(t)$. Hence, $F^*(-f)\, H_{eqz}(f) = 1$, where $F^*(-f)$ and $H_{eqz}(f)$ are the Fourier transform of the function $f^*(t)$ and $h_{eqz}(t)$ respectively.

If the channel is frequency selective, the equalizer enhances the frequency components with small amplitudes and attenuates the strong frequencies in the received frequency response. For a time-varying channel, an adaptive equalizer is needed to track the channel variations.

6.3 EQUALIZER NOISE ENHANCEMENT

Equalization is a process through which effect of ISI is mitigated. But it has to do in such a way that during this process noise power should not be enhanced. Figure 6.3 shows a typical analog equalizer. A signal $s(t)$ is passed through a channel with the response $H(f)$, at the receiver a noise $n(t)$ of type Gaussian is added. Output of the receiver $y(t) = s(t) + n(t)$ and in frequency domain $Y(f) = S(f)H(f) + N(f)$, where $N(f)$ is the noise power with spectral density N_0. Suppose, we want to equalize the ISI effect completely by introducing an equalizer with response function $H_{eqz}(f)$ such that $H_{eqz}(f) = 1/H(f)$.

The equalizer output is $s(t) + n'(t)$ and $Y'(f) = [S(f)H(f) + N(f)]\, H_{eqz}(f) = S(f) + N'(f)$, the equalizer output and the new Gaussian noise spectral density becomes $N_0/|H(f)|^2$.

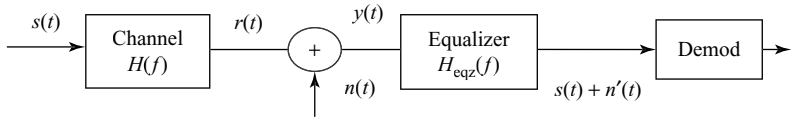


Fig. 6.3 *Analog equalizer illustrating noise enhancement*

At some frequencies if $H(f)$ is greatly attenuated then the equalizer will enhance the noise power thus reducing signal to noise ratio (SNR) at those frequencies as $H(f)$ is inversely related with $H_{eqz}(f)$. The effect of nullifying ISI would not be useful unless the SNR is maximized after equalization. Digital linear equalizer follows the inversion of frequency response thus enhances the noise power whereas nonlinear equalizers do not invert and less suffers from noise enhancement.

Example 6.1 If for a channel with bandwidth B = 30 kHz, the channel response $H(f) = 1/|0.5f|$ for $|f| < B$, Noise PSD = $N_0/2$, then what would be the noise power before and after equalization.

Solution Before equalization noise power = N_0B. With equalization PSD = $N_0/|H(f)|^2 = N_0 \times (0.25\,|f|^2)$ for $|f| < B$. So, noise power is

$$N_0 \int_{-B}^{B} 0.25 f^2 df = 0.25 N_0 f^3/3 = 0.25 N_0 2B^3/3 = 0.17 N_0 B^3$$

Huge increase of noise power because of equalization process is occurred.

6.4 TYPES OF EQUALIZER

There are two basic divisions of equalizer, one is **linear** equalizer and the other is **non-linear** equalizer. These categories are determined from how the output of an adaptive equalizer is used for subsequent control (feedback) of the equalizer. Though the linear equalizer is the simplest to understand but as discussed earlier it suffers from noise power enhancement compared to the non-linear equalizer and is not suitable for wireless applications. As shown in Fig. 6.2, the decision maker for the analog signal $d'(t)$ determines the value of digital data bit being received and applies a slicing or thresholding nonlinear operation in order to determine the value of $d(t)$. If feedback is used to determine the value of $d(t)$, then the equalizer is non-linear otherwise linear. Linear and nonlinear equalizers are generally implemented by using **transversal or lattice structure**. Equalizers are further divided according to the filter **implementation structure, and adaptive algorithms used**. Decision Feedback Equalizer (DFE) is the most simple nonlinear equalizer type. But it is not suitable for low SNR cases; bits are erroneously detected leading to poor performance. **Maximum Likelihood Sequence Equalizer** (MLSE) is the optimal equalizer with exponential increase in complexity when length of the delay spread is increased. Equalizers are also classified as Symbol-to-Symbol (SBS) or as Sequence Estimators (SE). SBS removes ISI from each symbol and then detect each symbol individually. DFE and all other linear equalizers are of SBS type. SE detects sequence of symbols and ISI effect is the part of estimation process. MMSE is the optimum form of sequence detection. Transversal structure uses filter with N–1 delay elements and N taps with complex weights to tune, whereas the lattice structure uses complex recursive structure with better numerical stability, faster convergence and greater flexibility in changing their length. Fig. 6.4 shows the different types of equalizers showing a logical classification.

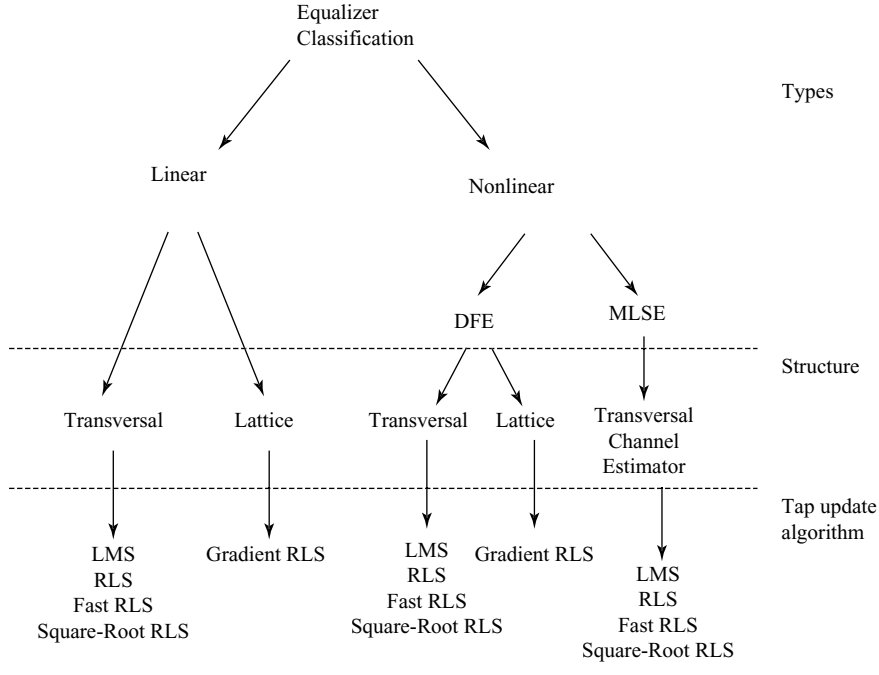


Fig. 6.4 *Classification of equalizers*

In wireless environment, channel is random and time variant. For a very known channel with static characteristics, **a filter can be designed for the equalization process**. By sending sequence of known **training pulses**, the receiver can estimate the channel impulse response $h(t)$ that corresponds to **minimum bit error rate** (BER) detection. Sending the training sequence (typically a pseudorandom binary signal or known bit

stream) repeatedly over a short interval of time solves the problem of time variance, so that the receiver **can adapt the situation of the channel**. That is why the concept is called **adaptive channel equalization**. In an adaptive equalizer, the receiver uses **recursive** algorithm **to estimate the channel characteristic** and **filter coefficients to minimize** the distortion that occurs due to multipath fading. The filter coefficients obtained in this manner is the optimum one and designed for the worst condition of the channel such as maximum time delay spread, fastest velocity and deep fading. **The convergence** is the main criterion in an adaptive equalization process for varied channel characteristics and **continually changing** filter coefficients.

Three factors affect the time spanning over which an equalizer converges: *equalizer algorithm, equalizer structure and time rate of change* of the multipath radio channel. Time division multiple access (TDMA) in wireless system is a well-suited system for equalizers.

6.4.1 Linear Transversal Equalizer

LTE is the simplest common equalizer that uses tapped delay lines as shown in Fig. 6.5(a).

The symbol period T_s is used to keep tap spacing. LTE filters uses only feed forward taps with delay operator z^{-1} (or exp $(-j\omega T_s)$. It has many poles or zeroes at z = 0, also known as ***Finite Impulse Response*** (FIR) filter. When both feed forward and backward taps are there, the filter is **Infinite Filter Response** (IFR) and its transfer function is a **rational function of z^{-1}**. The number of delay elements (the filter order) used in the filter, determines the finite duration of its filter response.

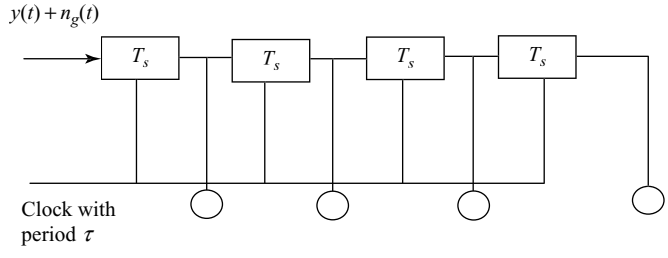


Fig. 6.5(a) *Linear transversal equalizer structure with feed forward delay*

The linear equalizer can be implemented by using transversal filter, which is a tapped delay line filter. The **current and past values of the received signals are weighted by the filter coefficient** and then added to produce output as shown in Fig. 6.5(b).

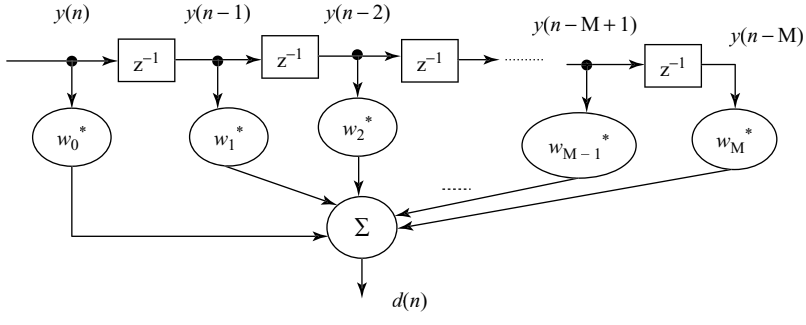


Fig. 6.5(b) *Structure of transversal filter, linear equalizer*

When z^{-1} operates on the input $y(n)$, the resulting output is $y\,(n-1)$. The role of the multiplier in the filter is to multiply tap input to which it is connected by the filter coefficient known as tap weight. Thus, a multiplier connected to the k^{th} tap input $y\,(n-k)$ produces $w_k^*\,y\,(n-k)$, where w_k is the respective tap weight and k = 0,1, ….M, where M is the filter order. The asterisk denotes the complex conjugation, which assumes that the

tap inputs and therefore the tap weights are all complex valued. The overall output of the filter is the added sum of the individual multiplier output.

$$d(n) = \sum_{k=0}^{M} w_k^* y(n-k) \tag{6.7}$$

The sum is known as the *finite convolution sum*. The length of the **equalizer M** is dependent on the implementation considerations. As M increases, complexity of the system will increase.

If we consider (2M +1) taps with coefficients w_{-M} to w_M, then the transfer function of the equalizer in the z-domain is given by

$$D(z) = \sum_{-M}^{M} w_k^* z^{-k} \tag{6.8}$$

Considering the equalized system of Fig. 6.6, The z transform of the discrete time sequence $\{r_k\}$ is R(z) and is given by Eq. (6.9)

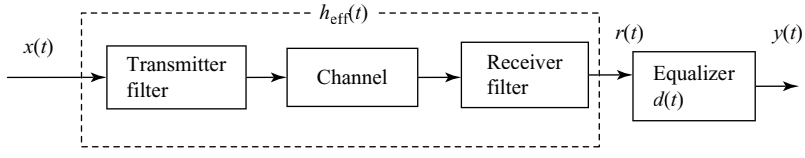


Fig. 6.6 *Equalized system*

$$R(z) = \sum_{-\infty}^{\infty} r_n z^{-n} \tag{6.9}$$

and the equalizer output in the z domain is

$$Y(z) = R(z)\, D(z) = \sum_{n=-\infty}^{\infty} r_n z^{-n} \sum_{k=-M}^{M} w_k z^{-k} = \sum_{n=-\infty}^{\infty} \sum_{k=-N}^{N} w_k r_n z^{-(k+n)}$$

Let $m = k + n$, then

$$Y(z) = \sum_{n=-\infty}^{\infty} \sum_{k=-N}^{N} w_k r_{m-k} z^{-m} = \sum_{m=-\infty}^{\infty} y_m z^{-m} \tag{6.10}$$

For **a given channel**, the optimal choice of equalizer coefficients would be the **coefficients that minimize probability of error**. But optimizing $\{w_k$'s$\}$ is difficult. Indirect optimization that balances ISI mitigation with the prevention of noise enhancement is applied. There are two types of linear equalizers: the **Zero Forcing** (ZF) equalizer and the **Minimum Mean Square Error** (MMSE) equalizer. ZF can mitigate ISI considerably but with noise enhancement. MMSE minimizes expected mean square error between the transmitted signal and the detected signal at the equalizer output providing a balance between the **ISI mitigation and noise enhancement**.

6.4.2 Zero Forcing (ZF) Equalizer

For zero forcing equalizer, the filter coefficients would be such that the **combined channel and equalizer impulse response** is forced to become zero at all points except one of the sampled point NT_s of the tapped

delay line filter. The number of filter coefficients can be increased to a high extent for infinite length filter for which ISI will be zero at the output. For this the time delay for each delay elements will be equal to **symbol duration T_s**. The frequency response of such an equalizer is periodic with a period equal to symbol rate $1/T_s$, because of the T_s-second tap spacing. After sampling the effect of the channel on the received signal is determined by the folded frequency response $H_{ch}(f)$. Let the equalizer response is $H_{eqz}(f)$. For zero forcing, the combined filter response of the channel must satisfy the zero ISI condition and needs to be satisfied the Nyquist's first criterion as given below,

$$H_{ch}(f)H_{eqz}(f) = 1 \text{ for } |f| \angle 1/2T_s \tag{6.11}$$

So, an infinite-length zero ISI equalizer is simply an inverse filter which inverts the folded frequency response of the channel. If the folded frequency response is greatly attenuated then because of the inverse relationship filter may enhance the noise at this frequency, which is undesirable and is subjected to frequency selective fading.

With respect to Fig. 6.7, the z domain analysis for zero forcing equalizer can be done as follows.

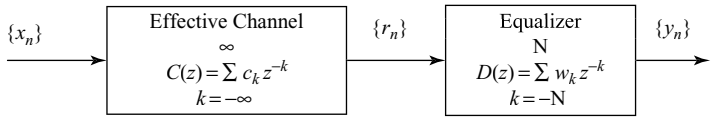


Fig. 6.7 *Discrete time representation of the zero forcing equalized system*

The transfer function of the equalized output $H(z) = Y(z)/X(z) = C(z)\,D(z)$

and in the time domain $h_n = c_{n^*} d_n = \sum_{k=-N}^{N} w_k c_{n-k}$ (6.12)

In the ISI free transmission, the output of the equalizer $Y(z)$ should be equal to $X(z)$, which requires that $H(z) = 1$, that implies the channel impulse response in the time domain h_n should satisfy the following condition,

$$h_n = \begin{cases} 1, & \text{for } n = 0 \\ 0, & \text{for } n \neq 0 \end{cases} \tag{6.22}$$

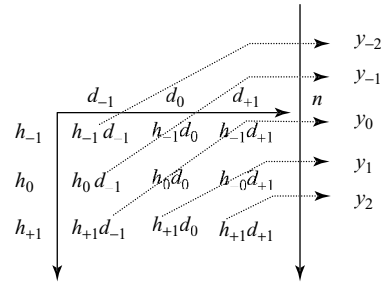
Example 6.2 Consider the discrete frequency channel response in a multipath wireless propagation is h(n) = [.2, .9, .4]. Design the 3-tap zero forcing equalizer for the system for ISI free transmission. Plot the frequency domain channel and equalization response. Also plot the frequency response curves for 5 tap and 7 tap equalizers.

Solution For 3-tap equalizer, let y_n be the equalizer output, then it is the convolution of d_n (equalizer response) and h_n (channel response)

As, $y_n = d_n * h_n$, where h_n is the channel response, d_n is the equalizer response.

Our requirement is to get values of the equalizer coefficients for 3-tap equalizer as d_{-1}, d_0 and d_{+1}

For 3-tap equalizer, y_n will have 5 samples which are the convolved output of d_n and h_n, n varies from –2 to +2 as shown in the following figure.



From the convolution equation, the following three conditions are obtained:

$$h_0 d_{-1} + h_{-1} d_0 = y_{-1} = 0$$
$$h_{+1} d_{-1} + h_0 d_0 + h_{-1} d_{+1} = y_0 = 1$$
$$h_{+1} d_0 + h_0 d_{+1} = y_1 = 0$$

In matrix form,

$$\begin{bmatrix} h_0 & h_{-1} & 0 \\ h_{+1} & h_0 & h_{-1} \\ 0 & h_{+1} & h_0 \end{bmatrix} \begin{bmatrix} d_{-1} \\ d_0 \\ d_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

In this problem, $h_0 = 0.9$, $h_{-1} = 0.2$ and $h+_1 = 0.4$. After matrix inversion, we get the equalizer coefficients as

$$d_0 = 1.3846,\ d_{-1} = -0.3077 \text{ and } d+_1 = -0.6154$$

Now the channel frequency response is $H(e^{jw}) = 0.2\, e^{jw} + 0.9 + 0.4\, e^{-jw}$

$$= (0.6 \cos w + 0.9) - j\,(0.2 \sin w)$$

The equalizer frequency response is $D(e^{jw}) = d_{-1}\, e^{jw} + d_0 + d_1\, e^{-jw}$

$$= (1.3846 - 0.921 \cos w) + j\,(0.3077 \sin w)$$

The design of 5-tap and 7- tap equalizer can also be found in a similar manner. Using the MATLAB program the plots for frequency response for 3, 5 and 7 tap equalizers are given in the Fig. 6.8.

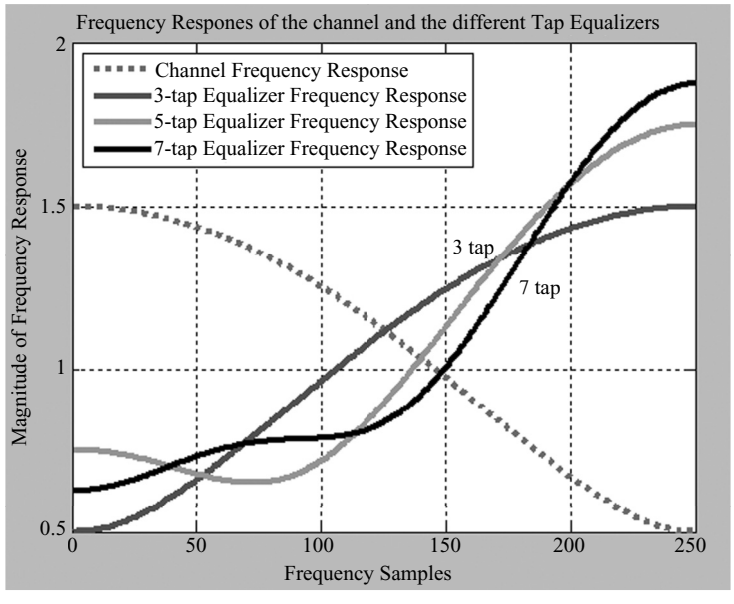


Fig. 6.8 *Channel frequency response and equalizer frequency response curves for ZF equalizer with different taps*

Example 6.3 Analysis of bit error probabilities for the 3, 5, 7 taps equalizers. Consider the effective channel impulse response $h(n) = [1, .5] = \delta(n) + 0.5\, \delta(n-1)$. Considering the problem of zero forcing equalization for ISI free transmission using 3, 5 and 7 tap equalizers, analyze the bit error probabilities and give the MATLAB plots for these.

Solution The z-domain transfer function $H(z) = 1 + 0.5\, z^{-1}$. Consider $D(z)$ be the equalizer response, then in ideal case $D(z) = 1/H(z) = 1 / (1 + 0.5\, z^{-1}) = 2z / 2z + 1$

Assuming $D(z)$ to be causal system, we write $d(n) = (-1/2)^n\, u(n)$, where ROC : $|z|>1/2$

This implies $d(n) = [\, 1, -1/2, +1/4, -1/8, +1/16, \ldots..]$
Considering 3-tap equalizer, $d(n) = [d_{-1} = 0, d_0 = 1$ and $d_1 = -0.5]$
For 5-tap equalizer the $d(n) = [\, d_{-2} = 0, d_{-1} = 0, d_0 = 1, d_1 = -0.5, d_2 = 0.25]$
For 7-tap equalizer $d(n) = [\, d_{-3} = 0, d_{-2} = 0, d_{-1} = 0, d_0 = 1, d_1 = -0.5, d_2 = 0.25, d_3 = -0.125]$
In the presence of some noise entering into the equalizer, the auto-correlation of the noise $R(k_1) = \sigma_v^2\, \delta(k_1)$, where $\sigma_v^2 = N_0/2$, the Noise Spectral Density.

Let v_n be the noise output of an infinite tap equalizer, then $v_n = \sum_{k=0}^{\infty} d_k\, n_{n-k}$, where n_n be the noise input at the equalizer. Now let us calculate the auto correlation of the output of the equalizer when input is AWGN with variance $N_0/2$.

$$R_{\text{eqzop}}(\tau) = \text{E}[V_{n+\tau} V_n^*], \text{ where } V_{n+\tau} = \sum_{k=0}^{\infty} d_k n_{n+\tau-k}$$

$E\,[v_{n+\tau}\, v_n^*] = \text{E}\,[\sum_{p=0}^{\infty} d_p\, n_n + \tau - p. \sum_{q=0}^{\infty} d_q\, n^*_{n-q}] = \sum_{p=0}^{\infty} d_p \sum_{q=0}^{\infty} d_q\, \text{E}\,[n_{n+\tau-p}\, n^*_{n-q}]$ (as all the equalizer coefficients are real).

Now the noise variance can be calculated by putting $\tau = 0$.

$$E[v_n v_n^*] = \sum_{p=0}^{\infty} d_p \sum_{q=0}^{\infty} d_q \text{E}[n_{n-p} n^*_{n-q}] = \sum_{p=0}^{\infty} d_p \sum_{q=0}^{\infty} d_q N_0 / 2\delta(p-q)$$

$$= \sum_{p=0}^{\infty} N_0 / 2.(d_p)^2 = N_0 / 2[\sum_{p=0}^{\infty} (1/4)^p] \text{ as } d(n) = (-1/2)^n\, u(n)$$

For infinite tap equalizer, $E\,[\, v_n\, v_n^*] = \sigma_v^2 = N_0/2\ \ [\, 1/ (1 - 1/4)] = 2\, N_0/3$
Thus the probability of bit error $= P_{\text{binfinite}} = Q(\sqrt{E_b / 2N_0 / 3}) = Q(\sqrt{3E_b / 2N_0})$

For 3-tap equalizer $\sigma_v^2 = (N_0/2\,)\,[1 + 1/4] = 5N_0/8$

$$P_{\text{b-3tap}} = \text{Q}(\sqrt{8E_b / 5N_0})$$

For 5-tap equalizer, $\sigma_v^2 = (N_0/2)\ \ [1 + 1/4 + 1/16] = 21\, N_0/\, 32$

$$P_{\text{b-5tap}} = Q(\sqrt{32E_b / 21N_0})$$

For 7- tap equalizer, $\sigma_v^2 = (N_0/2)\ \ [1 + 1/4 + 1/16 + 1/64] = 85\, N_0/\, 128$

$$P_{\text{b-7tap}} = Q(\sqrt{128E_b / 85N_0})$$

Since Q is a decreasing function of its corresponding argument,

$$P_{\text{b-7tap}} > P_{\text{b-5tap}} > P_{\text{b-3tap}}$$

The MATLAB plot for error probability for infinite, 3, 5 and 7 tap equalizers is shown in Fig. 6.9.

It is seen from Fig. 6.8 that the equalizer frequency response is in good agreement with the inverse of channel frequency response for 3-tap equalizer rather than the 5 and 7 taps. This has happened because in ZF equalizer, the probability of error increases with the number of taps and noise enhancement is also increased.

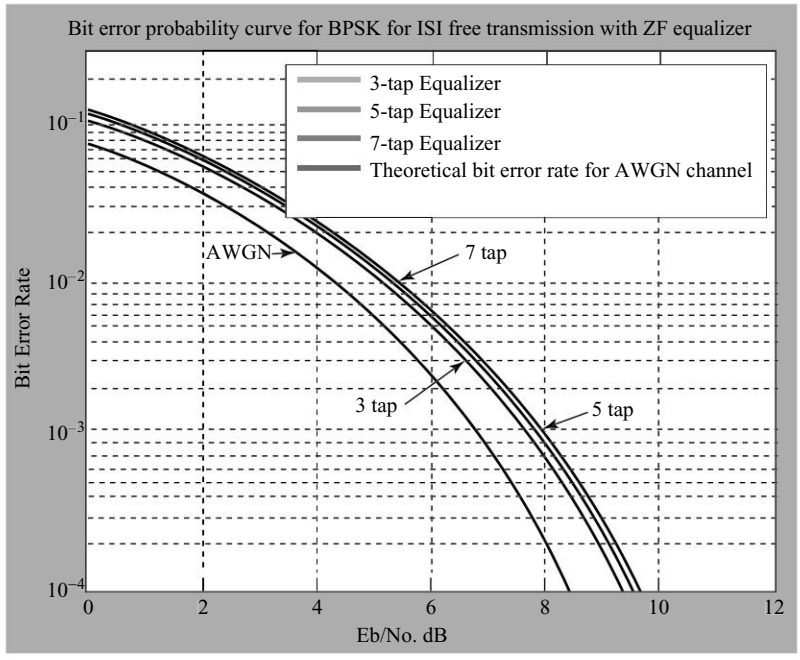


Fig. 6.9 *The bit error probability for ZF equalizer with different taps as calculated in Problem 6.2 for channel response h(n) = [0 1.0 .5]*

6.4.3 Minimum (Least) Mean Square Equalizer

ZF equalizer enhances the noise power and increases bit error probability. In case of frequency selective channel, there is a chance of having high attenuation in some frequency than the other. As the effective channel has high attenuation, the equalizer can significantly enhance noise in this frequency, as it has inverse channel response. As a result transmission performance degradation is considered in designing the equalizer both for ISI and channel noise. This effect can be minimized using Least (Minimum) Mean Square Equalizer (LMS). LMS equalizer is a very robust one where minimization of the mean square error (MSE) between the desired equalizer output and the actual equalizer output is obtained. LMS equalizer maximizes the signal to distortion ratio at its output within the constraints of the equalizer time span and the delay through the equalizer, which is dependent on the position of the tap. Let d_k is the transmitted signal and its estimate is d_k', then MMSE is a linear equalizer in which the weights $\{w_i\}$'s are chosen so that $E[d_k - d_k']^2$ is minimized. The estimated output d_k' is the linear combination of the input samples $y[k]$:

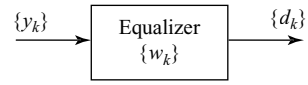
$$d_k' = \sum_{-M}^{M} w_i \; y\,[k - i] \tag{6.23}$$

Finding the **optimal filter coefficients** $\{w_i\}$ becomes the objective of the linear estimation in terms of MSE. If the noise input to the equalizer is of type white, then solution comes from **Weiner filtering** problem where concept of whitening filter is applied for the equalizer filter response. For a specific channel condition, the predicted error is dependent on the tap gain vector w_i. Since the channel characteristics are changing with time, it is difficult to estimate the receiver response every time. The changing scenario is considered with the help of adaptive equalizer in which the error signal e_k is fed back to adjust the filter weights until the error gets minimized.

In adaptive filter algorithm, there are two important processes, **training** and **tracking**. During the training period, a known signal is transmitted and a synchronized version of this sequence is generated at the receiver to estimate the channel condition. This training sequence may be the periodic isolated pulses and known maximal length PN sequences. The length of the training sequence should at least be equal to the length of the equalizer so that the transmitted signal spectrum is sufficiently dense in the channel bandwidth to be equalized. By knowing the synchronized version of the training signal, the error signal e_k is adjusted in mean square sense by controlling the equalizer coefficients. Updating the equalizer tap's gain, during each symbol duration does this adjustment process. In the mean square error, the MSE is the quadratic function of the coefficients. The predicted error e_k is dependent on tap gain vector w_i for a specified channel condition and the mean square error (MSE) of the equalizer is the function of w_i. If $J(w_i)$ denotes the cost function for MSE as a function of tap gain w_i, then the condition for MSE would be obtained by finding the derivative of $\xi = E(e_k e_k^*)$ and equating it to zero, where e_k = difference between the desired output and the actual output $= x_k - d_k = x_k - w_k^T y_k$ with respect to Fig. 6.10.

The output of the equalizer sampled at t = kT is

$$d_k = \sum_{n=-N}^{N} w_n y_{k-n}$$



and the desired output of the equalizer at $t = nT$ is x_k. The equalization error $e_k = d_k - x_k$. The problem for MSE is to find the equalizer tap coefficients to solve the equation $\min_w E[e_k^2]$, where $w = (w_{-N}, \ldots, w_{-1}, w_0, w_1, \ldots, w_N)^T$. The MSE of the equalization is defined as,

$$\begin{aligned}\xi = E(e_k)^2 &= E[(x_k - \sum_{n=-N}^{N} w_n Y_{k-n})^2] \\ &= E[x_k^2] + \sum_{m=-N}^{N}\sum_{n=-N}^{N} w_n w_m E[(y_{k-n})(y_{k-m}) - 2\sum_{m=-N}^{N} w_m E[(x_k y_{k-n})] \\ &= E[x_k^2] + \sum_{m=-N}^{N}\sum_{n=-N}^{N} w_n w_m R_y(n-m) - 2\sum_{m=-N}^{N} w_m R_{xy}(n)\end{aligned} \quad (6.24)$$

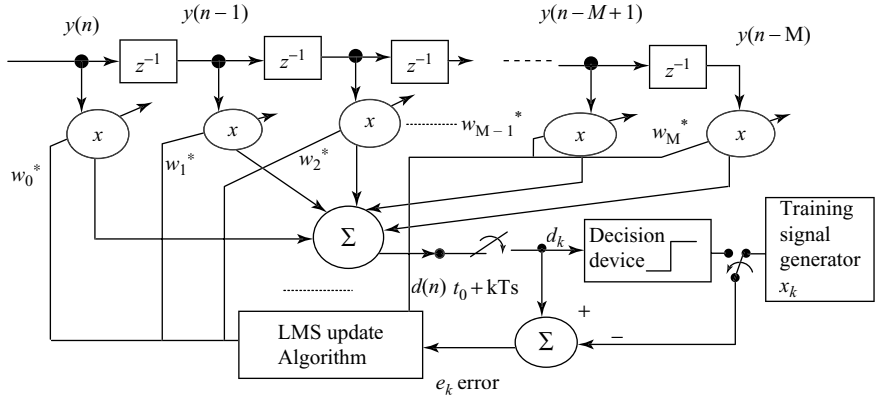


Fig. 6.10 *Adaptive equalizer using LMS*

The minimum MSE is obtained by differentiating ξ with respect to equalizer coefficients $\{w_n\}$, for $n = 0, \pm1, \pm2, \ldots \pm N$.

Thus, $\partial E(e_k^2) / \partial W_n(k) = 0$

$$\partial\xi / \partial W_n(k) = \sum_{m=-N}^{N}\sum_{n=-N}^{N} w_n w_m R_y(n-m) - 2\sum_{m=-N}^{N} w_m R_{xy}(n) = 0$$

$$\Rightarrow \sum_{m=-N}^{N} w_m R_y(n-m) = R_{xy}(n), n = 0, \pm 1, \pm 2, \ldots \pm \mathrm{N} \tag{6.25}$$

In matrix form, $R_y\ w = R_{xy} \Rightarrow w = R_y^{-1}\ R_{xy}$ (6.26)

Equation (6.26) is known as the *Wiener–Hopf* equation. R_y is the auto-correlation matrix and R_{xy} is the cross-correlation matrix, which are unknown *a priori*.

$$R_y = \begin{bmatrix} R_y(0) \ldots R_y(\mathrm{N}) \ldots\ldots R_y(2\mathrm{N}) \\ R_y(-1) \ldots\ldots R_y(\mathrm{N}-1) \ldots R_y(2\mathrm{N}-1) \\ \ldots\ldots \quad \ldots\ldots \quad \ldots\ldots \quad \ldots\ldots \\ R_y(-2\mathrm{N}+1) \ldots R_y(-\mathrm{N}+1) \ldots\ldots R_y(1) \\ R_y(-2\mathrm{N}) \ldots\ldots R_y(-\mathrm{N}) \ldots\ldots R_y(0) \end{bmatrix} \text{ and}$$

$$R_{xy} = [R_{xy}(-\mathrm{N}) \ldots \mathrm{R}_{xy}(-1) \ldots \mathrm{R}_{xy}(0) \ldots \mathrm{R}_{xy}(1) \ldots \mathrm{R}_{xy}(\mathrm{N})]^{\mathrm{T}}$$

Unlike the zero forcing equalizer, in the case of MMSE, the tap coefficients depend on the statistical properties such as correlation matrix and the vector of the noise as well as ISI. For time invariant channel, the auto-correlation matrix and cross-correlation matrix are obtained by transmitting a known training signal over the channel and time average estimation is taken.

Adaptive Linear Equalizer In a very fading dispersive channel, the channel response varies randomly with time. In such cases, the equalizer has to be adaptive to combat ISI effect by adjusting equalizer coefficients on line with some adaptive algorithm to adapt channel variations. When the equalizer is linear adaptive, the tap coefficients are such that it minimizes mean square error $\xi = E[e_k^2]$.

Using equation $e_k = x_k - d_k$ and $\min_{\mathrm{w}} E[e_k^2]$,

$$\partial\xi/\partial w_n(k) = E[2e_k \partial e_k / \partial w_n] = -2R_{ey}(k), \text{ where } R_{ey}(k) = E[e_k y_{k-n}]$$

If the channel variation rate is much smaller than the transmitted symbol rate, the optimal tap coefficients can be obtained by updating the tap gain coefficient iteratively according to

$$\begin{aligned} w_n(k+1) &= w_n(k) + \Delta[w_n(k)] \\ &= w_n(k) + \Delta\ E[\ e_k\ y_{k-n}] \end{aligned} \tag{6.27}$$

where n is the number tap delay in the equalizer (–N, –N+1,…–1, 0, 1…N), $w_n(k)$ is the n^{th} tap gain at time k, Δ is the positive adaptation factor or step size (sometimes used as μ), e_k is the error. $w_n(k)$ is the estimated w value at time t = kT, and $\Delta w_n(k)$ is the correction factor used to update new value for $w_n(k+1)$. The adaptive algorithm generates this correction factor based on input and error signal. There are two main adaptation algorithms one is the least mean square (LMS) and other is the recursive LMS. When the n-th element of $w(k) = w_n(k)$, $n = -\mathrm{N}, -\mathrm{N}+1, \ldots\mathrm{N}$ is not optimum then $\partial\xi / \partial \mathrm{w}_{\mathrm{n}}(k)$ is non zero. On the other hand, as $w_n(k)$ has reached the optimum value of w_n, the gradient $\partial e_k/\partial w_n$ and hence $R_{ey}(k)$ approaches zero.

Widrow and Hoff first introduced the LMS algorithm in 1959 as an adaptive algorithm, which uses a gradient-based method of steepest decent. LMS algorithm is used to minimize the error in least square sense by finding the suitable filter coefficient that relate to producing the LMS of the error signal (difference between the desired and the actual signal). It is a stochastic gradient decent method such that the filter is only adapted based on the error at the current time. LMS filter is built around a transversal, i.e., tapped delay line structure. LMS algorithm uses iterative procedure to make successive correction of the weight in the direction of the negative of the gradient vector that helps to reach minimum mean square error. It is relatively simple compared to other recursive algorithms, as it does not require the correlation matrix and matrix inversion.

Using the LMS update, the weight update is done as

$$w(k+1) = w(k) + \mu y(k) e^*(k), \text{where } e^*(k) = [x^*(k) - y^h(k) w(k)] \tag{6.28}$$

The initial value of the weight is considered as $w(0)$ at $k = 0$. The successive corrections of the weight vector ultimately reached to a minimum value of the MSE. The convergence of LMS algorithm is dependent on the step size μ and lies between $0 < \mu < 1/\lambda_{max}$, where λ_{max} is the largest eigen value of the correlation matrix $R = y(n)^* y(n)^h$. If μ is chosen to a very small value then the algorithm converges very slowly. On the other hand, a large value of μ may lead to a faster convergence with less stability around the minimum value.

Three factors affect the time spanning over, which an equalizer converges: equalizer algorithm, equalizer structure and time rate of change of the multipath radio channel. For adaptation, a training sequence (normally pseudorandom sequence) is transmitted before the information data sequence to obtain the initial optimum tap coefficients of the adaptive equalizer. As the receiver knows the sequence, the tap coefficients are adjusted to optimum value until the known sequence is obtained. After this, the adaptive equalizer estimates the equalizer error for the detected symbols and the LMS algorithm is used to minimize the error by updating tap coefficients as described.

6.4.4 Decision Feedback Filter (DFE)

Equalizers specifically designed for multipath correction are often termed *echo-cancellers* or *deghosters*. They may require significantly longer filter spans than simple spectral equalizers, but the principles of operation are essentially the same. A decision feedback equalizer (DFE) is a non-linear equalizer useful when channel is affected with severe amplitude distortion. The equalized signal is the sum of the outputs of the forward and feedback signal of the equalizer as shown in Fig. 6.11. The forward part is similar to the linear transversal filter as discussed earlier and the decision made on the equalized signal is fed back through a second transversal filter. If the past decision is assumed to be correct for the detected signal, then the ISI for these symbols can be exactly cancelled by subtracting the past symbol values along with the use of appropriate weight from the equalizer output. The forward and feedback coefficients may be adjusted simultaneously to minimize the MSE.

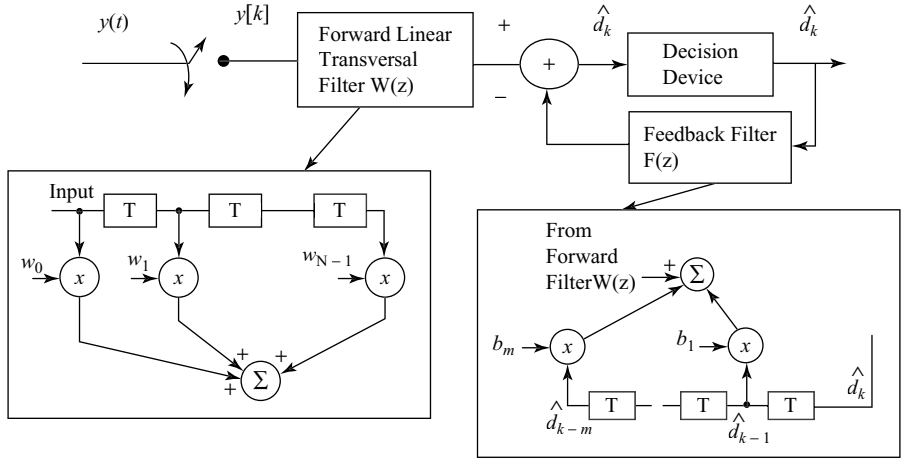


Fig. 6.11 *Decision feedback filter structure*

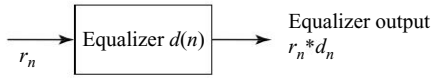
The update for the forward filter coefficients is same as for the linear equalizer and the feedback coefficients are adjusted according to the equation,

$$b_m(k+1) = b_m(k) + \Delta e_k \hat{d}'_{k-m}, \tag{6.29}$$

where $m = 1, 2, \ldots.M$, and $\hat{d'}_k$ is the k^{th} symbol decision, $b_m(k)$ is the m^{th} feedback coefficient at time k. There are M feedback coefficients in all. For optimum values of b_m, ISI will be zero within the span of feedback part as in ZF equalizer. The feedback coefficients has no effect in determining the equalizer output as the output of the DFE equalizer is the weighted sum of the noise free past decision [1]. DFE is useful for channels with deep spectral nulls than the linear equalizer.

Example 6.4 Find the equalizer coefficients using 3-tap delay channel with response $h(n) = [h_{-1} = 0.2, h_0 = 0.9$ and $h_1 = 0.4]$ and $h(n) = [0\ 1.0\ .5]$ using minimum mean square error (MMSE) algorithm.

Solution Let the equalizer input in presence of AWGN is $r_n = y_n + w_n$ where w_n is the noise part.



Let s_n be the transmitted symbols such that $s_n \in R$ and $E[s_n s_k] = \delta(n - k)$. Also $E[s_n] = 0\ \forall\ n$. All the symbols are independent and identically distributed.

The signal received at the input of the equalizer is $r_n = y_n + w_n = \sum_{-1}^{1} h_k s_{n-k} + w_n$

Consider 3-tap equalizer with coefficients d_{-1}, d_0 and d_1.

The estimate of $s_n = \hat{s}_n = \sum_{-1}^{1} d_k r_{n-k}$

So the error is $e_k = s_n - \hat{s}_n$ and $\xi = E[e_k^2] = E[(s_n - \hat{s}_n)^2]$

The goal of MMSE equalizer is to minimize ξ.

Differentiating w.r.t d_q, $\partial\xi / \partial d_q = E[\partial/\partial d_q (s_n - \hat{s}_n)^2]$, where $q \in [-1, 0, 1]$

$$= E[\partial/\partial d_q (s_n - \sum_{-1}^{1} d_k r_{n-k})^2]$$

For MSE $\partial\xi / \partial d_q = 0$

$$\partial\xi / \partial d_q = E[-2\, d_q r_{n-q} (s_n - \sum_{-1}^{1} d_k r_{n-k})] = 0$$

as $d_q \neq 0$ for $\forall$ q, so

$$E[r_{n-q} s_n - \sum_{-1}^{1} d_k r_{n-k} r_{n-q})] = 0$$

$$E[r_{n-q} s_n - \sum_{-1}^{1} d_k E(r_{n-k} r_{n-q})] = 0$$

$$\sum_{-1}^{1} d_k R_{rr}(q-k) = R_{sr}(q)$$

$$\Rightarrow d_{-1} R_{rr}(q+1) + d_0 R_{rr}(q) + d_1 R_{rr}(q-1) = R_{sr}(q)$$

For $q = -1, 0, +1$ we get 3-equations

$d_{-1} R_{rr}(0) + d_0 R_{rr}(-1) + d_1 R_{rr}(-2) = R_{sr}(-1)$

$d_{-1} R_{rr}(1) + d_0 R_{rr}(0) + d_1 R_{rr}(-1) = R_{sr}(0)$

$d_{-1} R_{rr}(2) + d_0 R_{rr}(1) + d_1 R_{rr}(0) = R_{sr}(1)$

We know that the auto correlation is conjugate symmetric,

$R_{rr}(p) = R_{rr}(-p)$ for real p.

In matrix form,

$$\begin{bmatrix} R_{rr}(0) & R_{rr}(1) & R_{rr}(2) \\ R_{rr}(0) & R_{rr}(1) & R_{rr}(2) \\ R_{rr}(0) & R_{rr}(1) & R_{rr}(2) \end{bmatrix} \begin{bmatrix} d_{-1} \\ d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} R_{sr}(-1) \\ R_{sr}(0) \\ R_{sr}(1) \end{bmatrix}$$

Now, $R_{rr}(0) = E[r_n\ r_n] = E[(y_n + w_n)\ (y_n + w_n)] = E[y_n\ y_n] + E[w_n\ w_n]$

(as signal and noise are uncorrelated)

$$= E\left[\sum_{-1}^{1} h_k\ s_{n-k}\ \sum_{-1}^{1} h_k\ s_{n-k}\right] + \sigma_w^2$$

$$= E[(h_{-1}\ s_{n+1} + h_0\ s_n + h_1\ s_{n-1})(h_{-1}\ s_{n+1} + h_0\ s_n + h_1\ s_{n-1})] + \sigma_w^2$$

$$= h_{-1}^2 + h_0^2 + h_1^2 + \sigma_w^2$$

$$R_{rr}(1) = E[r_n\ r_{n-1}] = E[(h_0\ s_n + h_{-1}\ s_{n+1} + h_1\ s_{n-1})(h_{-1}\ s_n + h_0\ s_{n-1} + h_1\ s_{n-2})] = h_{-1}\ h_0 + h_0\ h_1$$

$$R_{rr}(2) = E[r_n\ r_{n-2}] = E[(h_0\ s_n + h_{-1}\ s_{n+1} + h_1\ s_{n-1})(h_{-1}\ s_{n-1} + h_0\ s_{n-2} + h_1\ s_{n-3})] = h_1\ h_{-1}$$

$$R_{rr}(3) = E[r_n\ r_{r-3}] = E[(h_{-1}\ s_{n+1} + h_0\ s_n + h_1\ s_{n-1})(h_{-1}\ s_{n-2} + h_0\ s_{n-3} + h_1\ s_{n-4})] = 0$$

Similarly, $R_{rr}(p) = 0\ \forall\ p \geq 3$

$$R_{sr}(-1) = E[s_{n-1}\ r_n] = E[s_{n-1}(y_n + w_n)] = E[s_{n-1}y_n] = E[(h_0\ s_n + h_{-1}\ s_{n+1} + h_1\ s_{n-1})](s_{n-1}) = h_1$$

$$R_{sr}(0) = E[s_n\ r_n] = E[s_n\ Y_n)] = E[s_n\ (h_0\ s_n + h_{-1}\ s_{n+1} + h_1\ s_{n-1})] = h_0$$

$$R_{sr}(1) = E[s_n\ r_{n-1}] = E[s_n\ Y_{n-1})] = E[S_n\ (h_{-1}\ s_n + h_{-0}\ s_{n-1} + h_1\ s_{n-2})] = h_{-1}$$

In matrix form,

$$\begin{bmatrix} h_{-1}^2 + h_0^2 + h_1^2 & h_{-1}h_0 + h_0\ h_1 & h_1\ h_{-1} \\ h_{-1}h_0 + h_0\ h_1 & h_{-1}^2 + h_0^2 + h_1^2 & h_{-1}h_0 + h_0\ h_1 \\ h_1\ h_{-1} & h_{-1}h_0 + h_0\ h_1 & h_{-1}^2 + h_0^2 + h_1^2 \end{bmatrix} + \begin{bmatrix} \sigma_w^2 & 0 & 0 \\ 0 & \sigma_w^2 & 0 \\ 0 & 0 & \sigma_w^2 \end{bmatrix} \begin{bmatrix} d_{-1} \\ d_0 \\ d_1 \end{bmatrix} = \begin{bmatrix} h_1 \\ h_0 \\ h_{-1} \end{bmatrix}$$

For 3-tap channel with $h_{-1} = 0.0$, $h_0 = 1$, $h_1 = 0.5$, and $\sigma_w^2 = 1$, the equalizer coefficients with MMSE algorithm is obtained as $d_{-1} = 0.1185$, $d_0 = 0.5584$ and $d_1 = -0.1672$ and for $h_{-1} = 0.2$, $h_0 = .9$, $h_1 = 0.4$, $d_{-1} = -.0631$, $d_0 = 0.9138$ and $d_1 = -.2469$. Note that for zero forcing equalizer with $h(n) = [.2\ .9\ .4]$, $d_{-1} = -.3077$, $d_0 = 1.3846$ and $d_1 = -.6154$ were obtained.

Example 6.5 Plot the frequency response for MMSE equalizer using 3-, 5- and 7- taps channel with response $h(n) = [0\ 1.0\ .5]$ and $h(n) = [.2\ .9\ .4]$. Also plot the bit error probability of transmission through these channels with MMSE equalization. Compare the error probability for 3-tap zero forcing equalizer and MMSE equalizer.

Comparing Figs. 6.13 and 6.14 it is seen that for bad channel condition of case $h(n) = [.2\ .9\ .4]$, the more signal to noise ratio is required to maintain the same order of probability of error (10^{-5}). It is also observed from Fig. 6.15 that error probability is less in MMSE equalizer than ZF, which is in the right direction of theoretical explanations for MMSE.

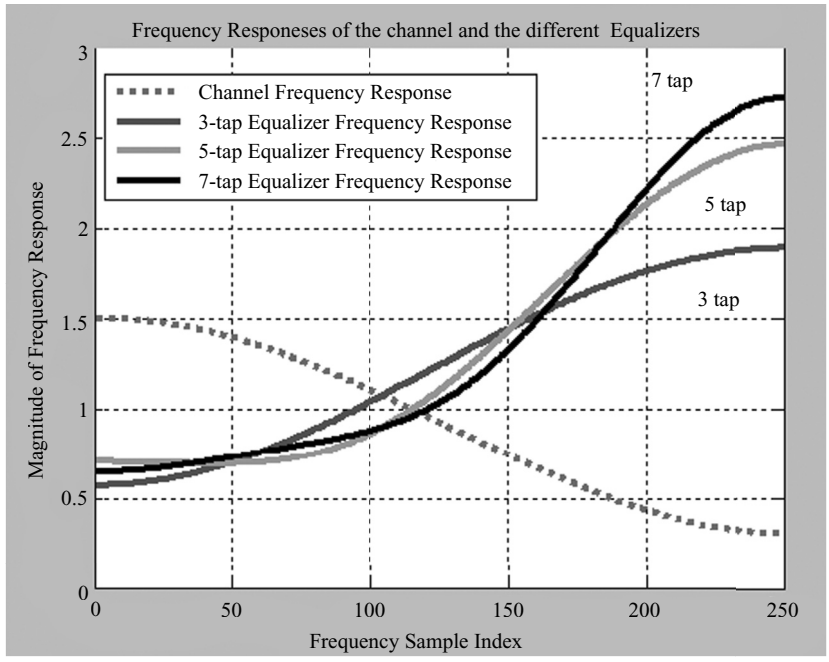


Fig. 6.12 *Frequency response of the channel and MMSE equalizer for different taps with h(n) = [0 1.0 .5]*

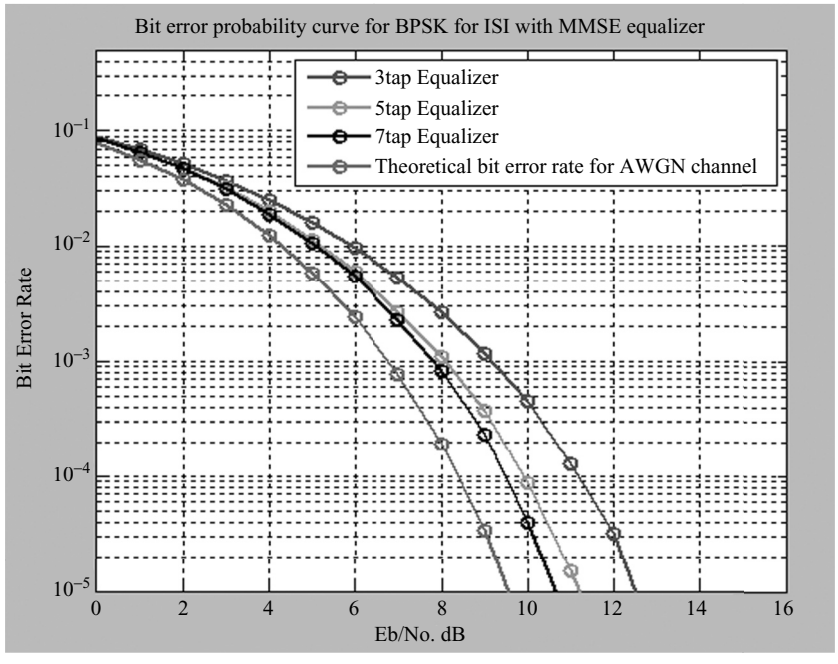


Fig. 6.13 *Bit error probability of MMSE equalizer with h(n) = [0 1.0 .5]*

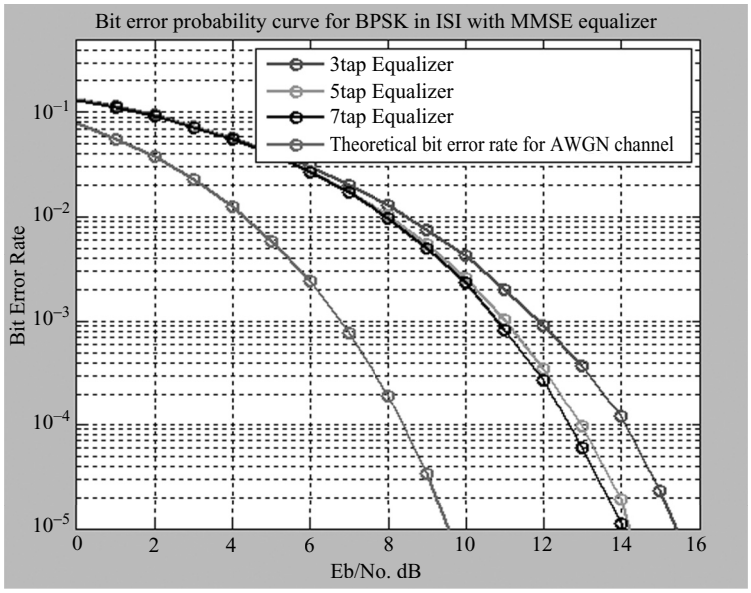


Fig. 6.14 *Bit error probability of MMSE equalizer for h(n) = [.2 .9 .4]*

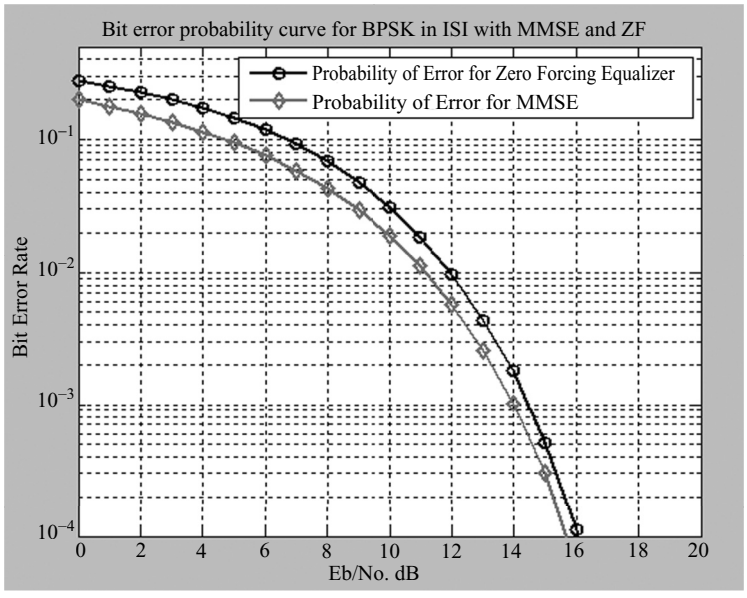


Fig. 6.15 *Comparison of bit error probability for ZF and MMSE with 3-tap equalizer*

6.5 DIVERSITY TECHNIQUES

Fading problem is a major impairment of the wireless communication channel. For Additive White Gaussian Channel (AWGN) the Bit Error Rate (BER) decreases exponentially with the Signal-to-Noise (SNR) ratio. A 10 dB SNR leads to BERs of the order of 10^{-4} whereas in a Rayleigh fading channel BER decreases linearly with the SNR. To obtain 10^{-4} BER it requires SNR of the order of 40 dB, which is very impractical. This happens because of the different fading occurs in different channel types. The trivial solution for the fading

problem would be to add a fading margin at the transmitter. However, this is not an efficient solution at all. One alternate solution is to take advantage of the statistical behavior of the fading channel. Here comes the basic concept of diversity; where two or more inputs at the receiver are used to get uncorrelated signals.

The principle of diversity is to ensure that the same information reaches the receiver on statistically independent channels. Diversity techniques can be used to improve system performance in fading channels. Instead of transmitting and receiving the desired signal through one channel, we obtain L copies of the desired signal through *M* different channels. The idea is that while some copies may undergo deep fades, others may not. We might still be able to obtain enough energy to make the correct decision on the transmitted symbol. There are several different kinds of diversity, which are commonly employed in wireless communication systems.

Diversity is most efficient when the different transmission channels or diversity branches carry independently fading copies of the same signal. Any correlation between the fading of the channels decreases the effectiveness of diversity. For any two signals x, y, the correlation coefficient is defined as,

$$\rho_{xy} = \frac{E\,[x.y] - E\,[x].E[y]}{\sqrt{E\,[x^2] - E\,[x]^2).(E[y]^2 - E\,[y]^2)}} \tag{6.30}$$

For any two statistically independent signal, $E[x.y] = E[x].E[y]$, and hence the correlation factor is zero.

Micro-diversity is the way of combating the small-scale fading that is differentiated into several categories. These are:

1. **Spatial diversity:** Several antennas are separated in space.
2. **Temporal diversity:** Repetition of transmitted signal occurs at different time.
3. **Frequency diversity:** Signals are transmitted at different frequencies.
   Angular diversity where multiple antennas with different patterns are used with or without spatial separation.
4. **Polarization diversity:** Multiple antennas are used for reception with different polarizations.

6.5.1 Frequency Diversity

In frequency diversity the same signal is transmitted at different frequencies as shown in Fig. 6.16. To achieve diversity, the information signal is to be modulated through *M* different carriers. Each carrier should be separated from the others by at least the coherence bandwidth B_c, so that different copies of the signal undergo independent fading. At the receiver, the *L* independently faded copies are "optimally" combined to give a statistic for decision. The optimal combiner is the ***maximum ratio combiner***, which will be introduced later. Frequency diversity can be used to combat frequency selective fading. Frequency diversity transmits information on more than one carrier frequency. Frequencies separated by more than the coherence bandwidth of the channel will not experience the same fades.

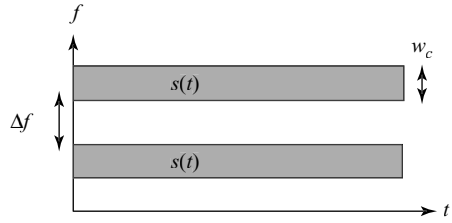


Fig. 6.16 *Frequency diversity*

6.5.2 Time Diversity/Temporal Diversity

Another approach to achieve diversity is to transmit the desired signal in *M* different periods of time, i.e., each symbol is transmitted *M* times. The intervals between transmissions of the same symbol should be at

least the coherence time T_c, so that different copies of the same symbol undergo independent fading. As the wireless propagation channel is time variant, signals those are received at different times are uncorrelated. For sufficient decorrelation, the temporal separation must be at least $1/2\gamma_{max}$, where γ_{max} is the maximum Doppler frequency shift.

Different ways of realizing temporal diversity are:

1. **Combination of interleaving and coding:** Error control coding, together with interleaving, can be an effective way to combat time selective (fast) fading. The different symbols of a codeword are transmitted at different times with the hope that at least one of them will reach with good SNR. The transmitted codeword then be recovered. But this type is only suitable for mobile environments.
2. **Automatic repeat request:** When transmission quality is bad due to fading, the receiver can send a message to transmitter with a request to resend the transmitted data after a wait of certain period that helps to achieve de-correlation. This method is bandwidth efficient.
3. **Repetition coding:** Transmission occurs many times for sending the same symbol whether repetition interval is large enough for decorrelation between the successive transmissions. This is very bandwidth inefficient.
   Optimal combining can also be obtained with the maximum ratio combiner.

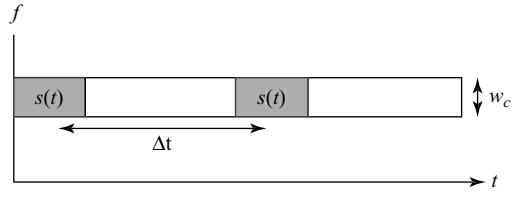


Fig. 6.17 *Temporal diversity*

Time diversity repeatedly transmits information at time spacing Δt that exceeds the coherence time of the channel as shown in Fig. 6.17.

6.5.3 Space Diversity/ Spatial Diversity

Spatial diversity is the oldest and simplest form of diversity. The transmit signal is received at several antenna elements. To achieve diversity, **M** antennas are used to receive **M** copies of the transmitted signal. The antennas should be spaced far enough apart so that different received copies of the signal undergo independent fading. The difference from frequency diversity and temporal diversity is that no additional work is required on the transmission end, and no additional bandwidth or transmission time is required. However, physical constraints may limit its applications. Sometimes, several transmission antennas are also employed to send out several copies of the transmitted signal. Spatial diversity can be employed to combat both frequency selective fading and time selective fading. The important design constraint for spatial diversity antenna is to establish the relationship between the antenna spacing and the correlation coefficient.

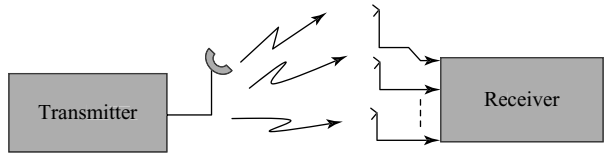


Fig. 6.18 *Selection diversity*

It is to be worth mentioning that in GSM cellular system at 900 MHz the minimum separation between antenna elements at the mobile station is only 8 cm, at 1800 MHz it would be 4 cm and for WLAN system at 2.4 GHz this separation is even smaller.

The idea of diversity is to combine several copies of the transmitted signal, which undergo independent fading, to increase the overall received power. The space diversity reception is again divided into following categories as:

1. Selection Diversity, 2. Maximal Ratio combining, and 3. Equal gain combining.

1. Selection diversity: In selection diversity method, the strongest signal branch is selected out of M signals as shown in Fig. 6.19.

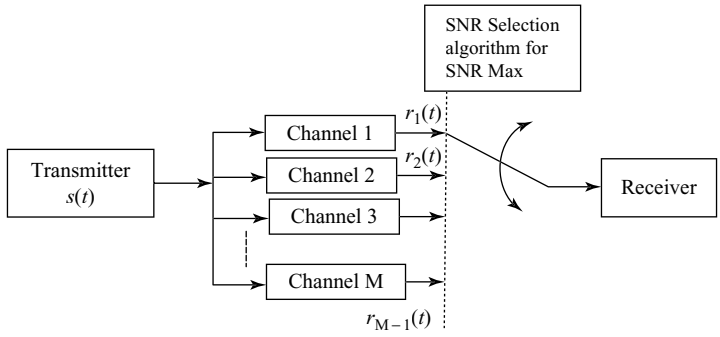


Fig. 6.19 *Selection diversity*

Out of M branches, M replicas of the transmitted signal are obtained as

$$r(t) = [r_1(t), r_2(t), \ldots r_{M-1}(t)]$$

where $r(t)$ is considered the sum of the received signal in slow fading channel and the additive Gaussian noise.

$$r_i(t) = s(t)A_i e^{j\theta i} + v_i(t), \qquad i = 0,1,2,\ldots\ldots M-1 \tag{6.31}$$

where $s(t)$ is the low pass equivalent of the transmitted signal, $A_i\, e^{j}\theta^{i}$ is the fading attenuation in channel i, $v_i(t)$ is the AWGN.

The output of the selection combiner is obtained for maximum of A_i and is given as

$$y(t) = Ae^{j\theta i} + s(t) + v(t), \text{ with } A = \text{Max}(A_0, A_1, A_2, \cdots A_{M-1}) \tag{6.32}$$

and at this condition the received signal strength is maximum given as $\text{SNR}_{\text{Max}} = \gamma_{\text{Max}} = A^2\, E_b / N_0 =$ Max ($\gamma_1, \gamma_2, \ldots\ldots\gamma_{M-1}$), where E_b/N_0 is the transmitted bit energy to noise ratio of the channel.

With uncorrelated branches of diversity channel and identically distributed (iid) fading, the pdf of the SNR in the diversity reception is given as

$$f_\gamma\,(x) = \begin{cases} 1/\gamma_0 \exp(-x/\gamma_0), & \text{if } x \geq 0 \\ 0, \text{ otherwise} \end{cases} \tag{6.33}$$

where $\gamma_0 = E(\gamma_m)$, γ_m is the received SNR per bit of the m-th channel at any instant.

With $x \geq 0$ and iid channel , cdf of the γ is

$$F_\gamma(x) = P(\gamma \le x) = P(\gamma_1 \le x \cap \gamma_2 \le x \cap \ldots\ldots \cap \gamma_m \le x) \tag{6.34}$$

$$= \prod_{m=1}^{M} P(\gamma_m \le x) = [\int_0^x f_\gamma(z)dz]^M \text{ with identically distributed fading}$$

Thus, $$P_\gamma(x) = [P_{\gamma_0}(x)]^M, \text{ and } P_\gamma(x) = MP_{\gamma_0}(x)[P_{\gamma_0}(x)]^{M-1} \tag{6.35}$$

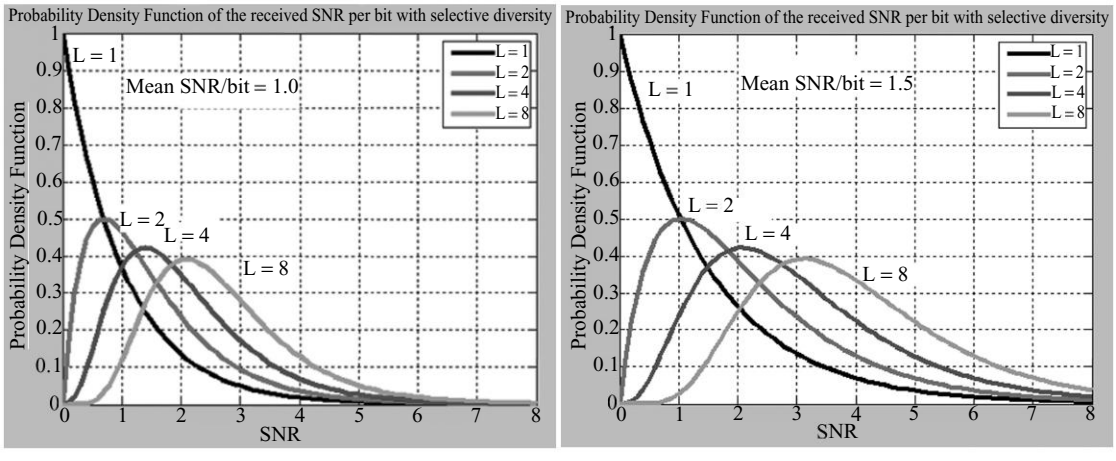
For Rayleigh fading channel, $P_\gamma(x) = (1 - e^{-x/\gamma_0})^M$, $\gamma_0 = 2\sigma^2 (E_b/N_0)$

The pdf of γ for Rayleigh channel

$$= f_\gamma(x) = dF_\gamma(x)/dx \tag{6.36}$$
$$= M(e^{-x/\gamma_0})\gamma_0(1 - e^{-x/\gamma_0})^{M-1}$$

Example 6.6 Consider L-th order diversity with selection combining in a Rayleigh fading propagation environment where each diversity channel exhibits independent and identically distributed fading. Plot the pdf of the received SNR per bit with selective diversity.

Solution As described in Section 6.5.3 for selection diversity we use Eq. (6.36) for the pdf of the SNR in Rayleigh fading channel and have written the MATLAB code, with varying number of channel L as 1, 2 , 3 and 4. We also consider γ_0 = mean SNR/bit to be equal to 1 and 1.5 and are plotted in Fig. 6.18(a) and (b) respectively. It is seen from Fig. 6.20, that for increased *L* from 1 to 8, chances of SNR <<1 is reduced more and more. The pdf curve shifts left to right that signifies that chance of having small value for the instantaneous SNR is less. This observation is more prominent when mean SNR/bit increases from 1 to 1.5 as with increased value of γ_0, chances of reception increases. Thus it can be conclude that with more diversity channels, selection would be better providing improved performance.



(a) (b)

Fig. 6.20 *Plot for pdf of the received SNR per bit with selective fading for different M for Rayleigh channel, (a) for mean SNR/bit (γ_0) =1.0 and (b) Mean SNR/bit =1.5*

2. Maximal ratio combining: Maximum ratio combining (MRC) is the optimum spatial diversity strategy to reduce the signal fluctuations caused by multipath propagation in wireless communications. By definition, a MRC combiner linearly combines the individually received branch signals so as to maximize the instantaneous output signal-to- noise ratio (SNR). In this method, the signals from all of the M diversity branches are weighted according to their individual signal power to noise power ratio for maximum SNR and then summed up as shown in Fig. 6.21. Unlike the selection diversity, the individual signals must be co-phased before summing up, which requires individual receiver and phasing circuit for each antenna element. The output SNR is the sum of the individual SNR. MRC technique provides the best statistical reduction of multipath fading and mostly used in modern communication systems.

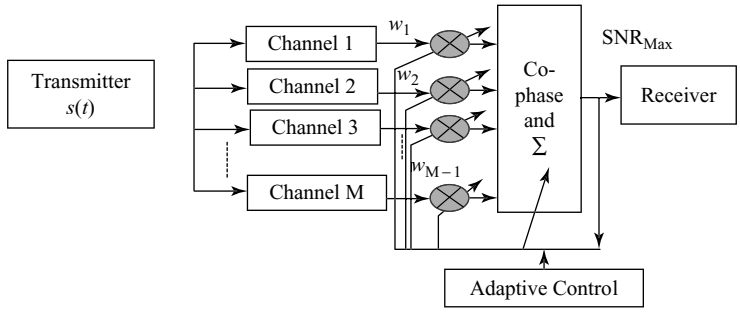


Fig. 6.21 *Maximal ratio combining*

The MRC output is given as

$$y(t) = \sum_{i=0}^{M-1} w_i \; r_i \; (t), \text{ where } w_i \; s \text{ are the channel gain conjugate,}$$

$$y(t) = \sum_{i=0}^{M-1} A_i \; e^{-j\theta i} \; r_i \; (t) = \sum_{i=0}^{M-1} A_i \; e^{-j\theta i} \; [\, A_i \; e^{j\theta i} \; s \; (t) + v_i \; (t)] \tag{6.37}$$

$$= \sum_{i=0}^{M-1} A_i^2 s \; (t) + \sum_{i=0}^{M-1} A_i \; e^{-j\theta i} v_i \; (t)$$

The SNR of the combined signal is, $\gamma_{\text{MAX}} = \sum_{i=0}^{M-1} A_i^2 E_b / N_0 = \sum_{i=0}^{M-1} \gamma_i$ (6.38)

where E_b/N_0 is the SNR for AWGN channel with $A_i = 1$ and $M = 1$. In a Rayleigh fading channel, the A_i *s* are iid Rayleigh random variables with parameters σ^2_{A}. The pdf of the channel for MRC is given by

$$f_y(x) = \frac{x^{\text{M}-1} e^{-\text{x}/\gamma_0}}{(\gamma_0)^M (M-1)!}, \text{ for x} \geq 0 \tag{6.39}$$

where $\gamma_0 = 2\sigma^2_A E_b / N_0$ the average SNR per bit in each diversity channel. The mean of the SNR per bit after combining is $\gamma_c = M \gamma_0$.

The outage probability is $P_{\gamma\text{out}}(x) = 1 - e^{-x/\gamma_0} \sum_{m=1}^{M} (x/\gamma_0)^{M-1}/(M-1)!$ (6.40)

Figure 6.22 is the plot for pdf of the received signal SNR per bit for maximal ratio combining .

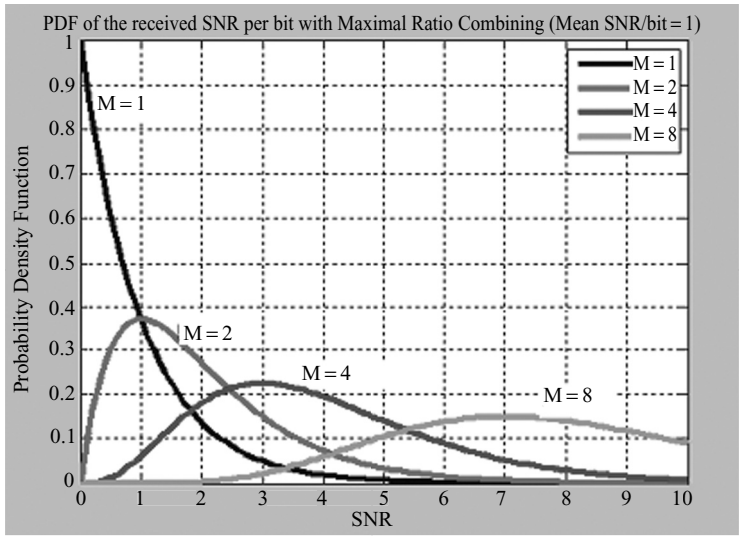


Fig. 6.22 (a) *Probability distribution function of the received signal SNR/bit for maximal ratio combining*

From Fig. 6.22(a), it is seen that as M increases the pdf curves shifts from left to right indicating that the chance for a small instantaneous SNR is small. Again comparing Fig. 6.22 with 6.21 for selection combining, it can be said that maximal ratio combining performs better to selection combining as the mean SNR per bit increases linearly with M for MRC.

3. Equal gain combining: In this case the branch weights are all set equal to unity with the signal co-phased to provide equal gain combining diversity. The performance of equal gain combining is better than selection diversity and inferior than maximal ratio combining.

Each of the branch signal is rotated by $e^{-j\theta i}$ and all branch signals are then added. The combiner output is given as,

$$y(t) = \sum_{i=0}^{M-1} e^{-j\theta i}\, r_i\,(t) = \left(\sum_{i=0}^{M-1} A_i \right) s\,(t) + \sum_{i=0}^{M-1} e^{-j\theta i}\, v_i(t) \tag{6.41}$$

The SNR is given by,

$$\gamma = \left(\sum_{i=0}^{M-1} A_i \right)^2 E_b \,/\, MN_0 \tag{6.42}$$

Diversity techniques are used to improve the performance of the radio channel without any increase in the transmitted power. As the received signal replicas have higher de-correlation, much diversity gain would be obtained. Among different combining techniques MRC has the best performance and the highest complexity, Selection diversity has the lowest performance and the least complexity. Equal Gain Combining performs very close to the MRC. Unlike the MRC, the estimate of the channel gain is not required in EGC.

Example 6.7 Find the probability of bit error for different modulation schemes of BPSK, QPSK and MSK using MCR diversity technique for Rayleigh fading channel.

Solution As discussed in Chapter 5, with coherent detection, BPSK, QPSK and MSK have the same bit error probability (BER) $= Q(\sqrt{2E_b/N_0})$, where E_b/N_0 is the signal to noise ratio per bit = γ_b

The probability of bit error over the Rayleigh fading channel is,

$$P_b = \int_{-\infty}^{\infty} p_e|_\gamma(x)\, f_\gamma(x)\, dx \tag{6.42a}$$

where $p_e|_\gamma$ is the conditional probability of bit error given that the received SNR per bit is $\gamma = x$. So, for coherent BPSK, QPSK and MSK, $p_e|(x) = Q(\sqrt{2x})$. Using Eqs.(6.42a) and (6.39), the bit error probability is

$$P_b = \int_0^{\infty} Q(\sqrt{2x}) \quad \frac{x^{M-1} e^{-x/\gamma_0}}{(\gamma_0)^M (M-1)!}\, dx = \frac{(1-\Gamma)}{2} \sum_{m=0}^{M-1} \binom{M-1+m}{m} \left(\frac{1+\Gamma}{2}\right)^m$$

where $\Gamma = \sqrt{\gamma_0/(1+\gamma_0)}$

With $\gamma_0 >> 1$, $(1+\Gamma)/2 \cong 1$ and $(1-\Gamma)/2 \cong 1/4\ \gamma_0$

and the error probability is approximated as,

$$P_b = (1/4\gamma_0)^M \binom{2M-1}{M}$$

Figure 6.22(b) shows the bit error rate of transmission performance using MCR diversity technique for different diversity branch M. It is observed that BER performance is much improved when M = 2 than M = 1 with respect to $\gamma_c = M\gamma_0$ in dB.

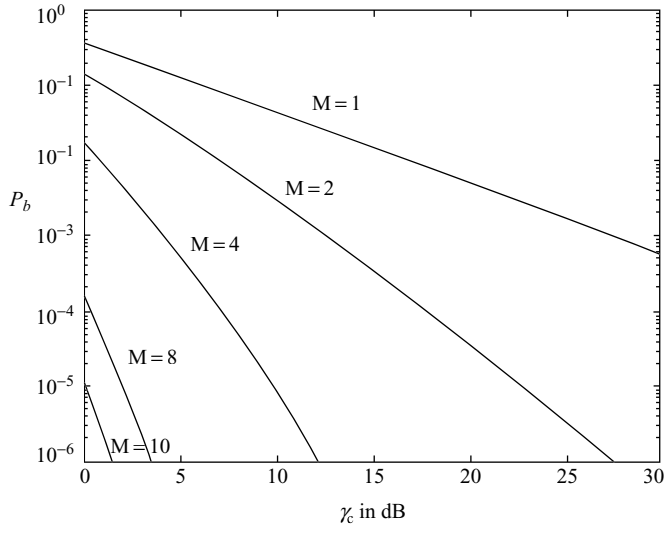


Fig. 6.22 (b) *BER performance for MRC diversity with i.i.d Rayleigh fading channel*

6.6 RAKE RECEIVER

In a mobile radio channel if reflected waves arrive with small relative time delays, self-interference occurs. Direct Sequence (DS) Spread Spectrum is often claimed to have particular properties that makes it less

vulnerable to multipath reception. In particular, the rake receiver architecture allows an optimal combining of energy received over paths with different delays. It avoids wave cancelation (fades) if delayed paths arrive with phase differences and appropriately weighs signals coming in with different signal-to-noise ratios. RAKE receiver, used specially in CDMA cellular systems, can combine multipath components to improve the signal to noise ratio (SNR) at the receiver, provides a separate correlation receiver for each of the multipath signals. Multipath components are practically uncorrelated when their relative propagation delay exceeds one chip period and equalization is not required. CDMA spreading codes are designed such that very low correlation exists between the successive chips. The basic idea of a RAKE receiver was first proposed by Price and Green and was patented in 1956.

The rake receiver consists of multiple correlators, in which the received signal is multiplied by time-shifted versions of a locally generated code sequence. The intention is to separate signals such that each finger only sees signals coming over a single (resolvable) path. The spreading code is chosen to have a very small autocorrelation value for any nonzero time offset. This avoids cross talk between fingers.

The rake receiver is designed to optimally detect a DS-CDMA signal transmitted over a dispersive multipath channel. It is an extension of the concept of the matched filter where the signal is correlated with a locally generated copy of the signal waveform. If, however, the signal is distorted by the channel, the receiver should correlate the incoming signal by a copy of the expected received signal, rather than by a copy of transmitted waveform. Thus the receiver should estimate the delay profile of channel, and adapt its locally generated copy according to this estimate.

6.6.1 Designing RAKE Receiver

Due to reflections from obstacles a radio channel can consist of many copies of originally transmitted signals having different amplitudes, phases, and delays. Multipath can occur in radio channel in various ways due to reflection, diffraction and scattering. The RAKE receiver uses a multipath diversity principle. It rakes the energy from the multipath propagated signal components as shown in Fig. 6.23.

Suppose that due to multipath propagation receiver gets M-ary multipath signals. An M-ary multipath model can be formed by considering each of the M paths with an independent delay t, and an independent complex time variant gain G to compensate the effect of different delays and phases.

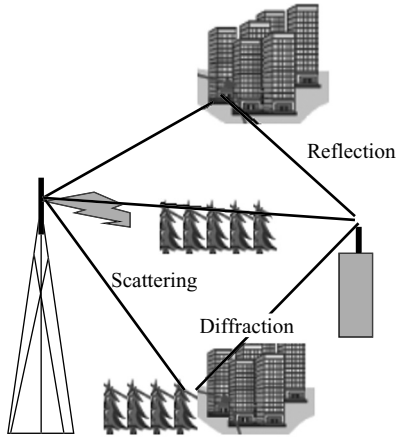


Fig. 6.23 *Multipath propagation*

Figure 6.24 is the simplified diagram for M-ary model for multipath propagation, where $s(t)$ is the transmitting signal, $r(t)$ is the received signal and $G(t)$ is the variable gain to combat the effect of different delay and phase.

Figure 6.25 shows the way to combine the multiple signals with the use of different co-rrelators in connection to DS-CDMA technology. RAKE receiver utilizes multiple correlators to separately detect M strongest multipath components. Each correlator detects a time-shifted version of the original transmission, and each finger correlates to a portion of the signal, which is delayed by at least one chip in time from the other fingers. A microprocessor controller can cause different correlation receivers to search in different time windows for significant multipath. The range of time delays that a particular correlator can search is called a *search window*. The RAKE receiver shown in Fig. 6.25 is essentially a diversity receiver designed specially for CDMA systems. The multipath components are uncorrelated to each other and relative delay spread between different components is greater than one chip rate.

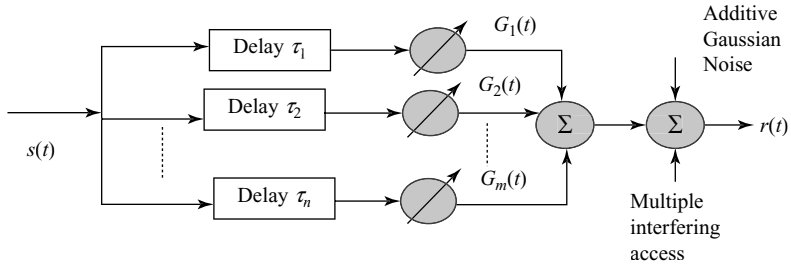


Fig. 6.24 *M-ary model for multipath propagation*

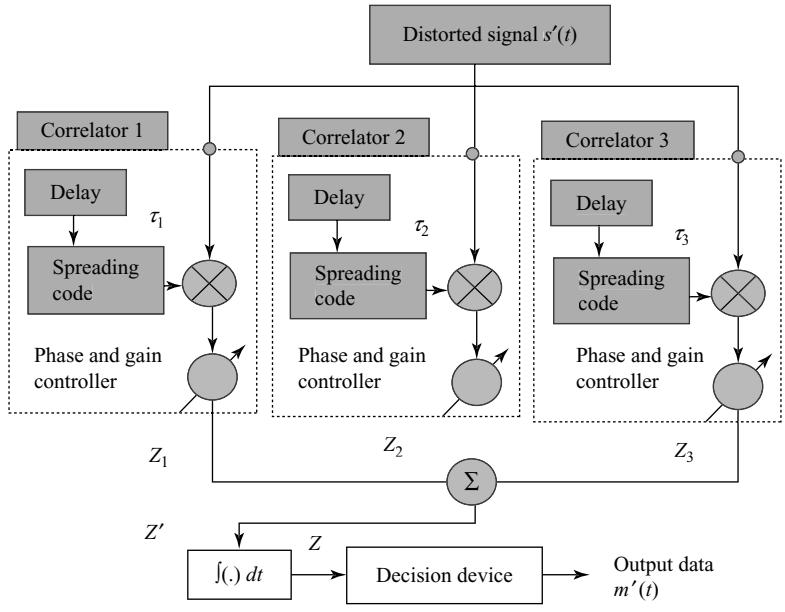


Fig. 6.25 *Technique to combat multipath effect in DS-CDMA system*

The output of each correlator in CDMA receiver is weighted to provide better estimate of the transmitted signal than is provided by a single component as shown in Fig. 6.26. Outputs of the M correlators are denoted as Z_1, Z_2,…, and Z_M. Correlator 1 is synchronized to the strongest multipath m_1 and correlator 2 is related to m_2, the signal m_2 arrives at a time τ_1 later than m_1, the difference of time between successive arrival $(\tau_2 - \tau_1)$ is less than one CDMA chip duration. The correlation between m_1 and m_2 is very low. If single receiver were used, then there may arise a situation when receiver cannot correct the received signal corrupted by fading.

Decision based on combination of *M* weighted outputs of correlators provide a low chance of error and creates a form of diversity technique in a fading situation and improve the reception quality in CDMA systems. The weighting factors are dependent on the power or signal to noise ratio at the correlator output. If SNR is small, then weight factor has small value.

Like a garden rake, the rake receiver gathers the energy received over the various delayed propagation paths. According to the maximum ratio combining principle, the SNR at the output is the sum of the SNRs in the individual branches, provided that, we assume that only AWGN is present (no other interference) and codes with a time offset are truly orthogonal. The output of the overall signal Z' is given by the summation as,

$$Z' = \sum_{m=1}^{M} \alpha_m \, Z_m \tag{6.43}$$

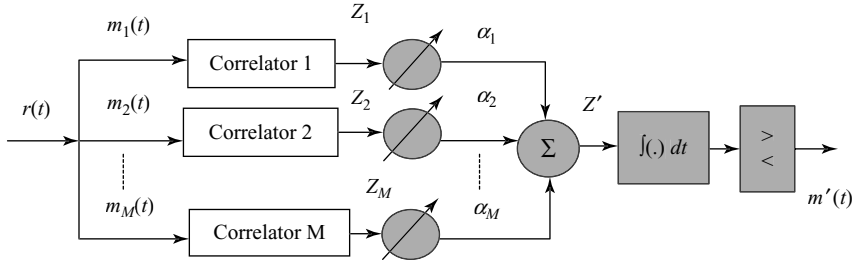


Fig. 6.26 *RAKE receiver in CDMA system*

The weighting coefficients, α_m are normalized to the output signal power of the correlator as,

$$\alpha_m = \frac{Z_m^2}{\sum_{m=1}^{M} Z_m^2} \tag{6.44}$$

There are many ways to generate the weighting coefficients. Due to Multiple Access Interference, RAKE fingers with strong multipath amplitudes will not necessarily provide strong output after correlation. Choosing weighting coefficients based on the actual outputs of the correlator yields to better RAKE performance.

6.7 CHANNEL CODING

Coding is used to detect or correct the erroneous bit of the transmitted signal in the wireless channel to be correctly decoded at the receiver. Channel coding generally used redundant bits to the transmitted signal. The codes uses for error detection are known as error detection code and the codes those are used for error correction are called error correction codes. Coding is different for AWGN channel and fading channel in wireless media. For AWGN channel the Shannon capacity is given by,

$$C = B \log_2\left(1 + \frac{P}{N_0 \, B}\right) = B \log_2\left(1 + \frac{S}{N}\right) \tag{6.45}$$

where B is the channel band width (in Hz), C is the channel capacity (bits/sec), P is the received power in watts = $E_b R_b$, N_0 is the single sided noise power density (Watts/Hz), E_b is the bit energy, R_b is the data (bit) transmission rate. The bandwidth efficiency is defined as

$$C/B = \log_2 (1 + E_b R_b / N_0 B) \tag{6.46}$$

Thus for higher signal-to-noise ratio (SNR), the bandwidth efficiency gets reduced because of the increase of the higher redundant bits to be used that increases raw data rate for a fixed rate source data transmission.

The main reason to apply coding in wireless channel is to reduce probability of error rate P_b during decoding process at the receiver. The block error rate or packet error rate (P_{bl}) is used in decoding the error within the block transmission, which is useful for accounting error in packet transmission. The probability of error reduction is dependent on coding gain in AWGN channel and diversity gain in fading channel. The coding gain in AWGN channel is defined as the amount of SNR that can be reduced by using coding technique for a specified P_b or P_{bl}.

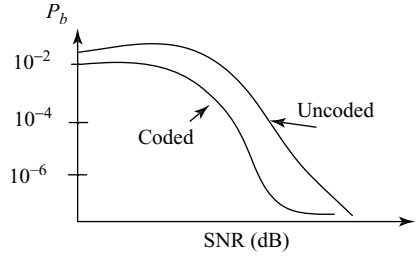


Fig. 6.27 *Typical relation of bit error probability vs. SNR*

Before transmission channel coder transform, the source code by adding code sequence and converted into coded signal. There are two types of error correction and detection codes namely Linear Block Codes and Convolutional Codes. Each codeword sent to the receiver may change during transmission. If the received codeword is the same as one of the valid code words, the word is accepted; the corresponding data word is extracted for use. If the received codeword is not valid, it is discarded. If corrupted but received word still matches, the error remains undetected. An error-detecting code can detect only the types of errors for which it is designed; other types of errors may remain undetected. An error correction is much more difficult than error detection.

6.7.1 Linear Block Code

It is used for forward error checking (FEC) with limited capacity of error detection and is corrected without retransmission. In the block codes, parity bits are added to blocks of source message bits and converted into code word or code blocks. It is very simple for implementation and is an extension of single parity bit check used for error detection. In a single parity bit, one extra bit is added in a block of n data bits to indicate whether the number of 1s in a block is odd or even. In case of occurring of single error, the parity bit may be corrupted and the number of 1s will be different in the transmitted bit sequence. So, a single bit error can be detected. Linear block codes extend this idea by using larger number of parity bits to either detect more than one error or to correct one or more bits. We will discuss about the binary codes where the data bits and coded word consisting of 0 or 1 only.

The receiver can detect a change in the original codeword if the following two conditions are met,

1. The receiver has or can find a list of valid code words.
2. The original codeword has changed to an invalid one.

The sender creates code words out of data words by using a generator that applies the rules and procedures of encoding.

Two important parameters for coding are the distance and weight of the codes. One of the central concepts in coding for error control is the idea of the Hamming Distance. It is the distance between two words (of the same size), determined by the number of differences between the corresponding bits. If two codes are C1 and C2, then the distance is denoted by d (C1, C2). The simple way to determine the Hamming Distance is to apply XOR operation on the two words and count the number of 1s in the result.

Example, HD d (000,011) is 2, because, 000⊕011 IS 011 (two 1s)

The number of non-zero elements in the codeword gives weight of a code, if binary code then the weight is basically the number of 1's in the codeword and is given by

$$w(C_i) = \sum_{j=1}^{N} C_{ij} \tag{6.47}$$

where j is the bit positional index and N is the length of the code word.

6.7.2 Binary Linear Block Codes

A desirable structure for a block code to possess is the linearity. With this structure, the encoding complexity will be greatly reduced. A binary block code generates a block of n coded bits from k number of information bits. The coded bits are called codeword. For all possible combination of n binary bits, the n codeword can take 2^n possible values, but only 2^k code words will be selected so that each k bit information block is uniquely mapped to one of these 2^k code words. The code rate is R_c (= k/n) information bits per codeword symbol. If R_s is the symbol transmission rate for the codeword, then the information rate associated with a (n, k) block code is $R_b = R_c R_s = (k/n) R_s$ bits/sec. Thus block coding reduces the data rate compared to the original un-coded transmission. A block code is called *linear*, when the mapping of the k information bits to the n codeword symbols is linear.

The binary information sequence is segmented into message block of fixed length, denoted by u. The encoder transforms each input message u into a binary n-tuple v with $n > k$.

This set of 2^k code words is called a block code. For a block code to be useful, there should be a one-to-one correspondence between a message u and its code word v. In fact, a binary block code is linear if and only if the module-2 sum of two code words is also a code word. Fig. 6.28 is the block diagram of Linear Block Code Generator.

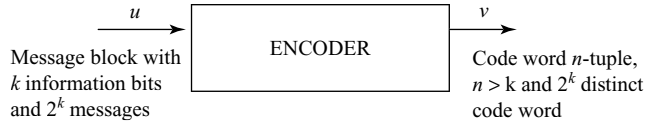


Fig. 6.28 *Block diagram of LBC*

Since an (n, k) linear code C is a k-dimensional subspace of the vector space V_n of all the binary n-tuple, it is possible to find k linearly independent code word, $\mathbf{g_0}$, $\mathbf{g_1}$, …, $\mathbf{g}_{k-1}$ in C. If $\mathbf{u} = (u_0, u_1, \ldots, u_{k-1})$ is the message to be encoded, the corresponding code word can be given as follows:

$$\mathbf{v} = u_0\, \mathbf{g}_0 + u_1\, \mathbf{g_1} + \ldots \ldots + u_{k-1}\, \mathbf{g_{k-1}} = \mathbf{u.\, G} \tag{6.48}$$

where, $u_i = 0$ or 1 for $0 \le i < k$

$$G = \begin{bmatrix} \mathbf{g_1} \\ \mathbf{g_2} \\ \vdots \\ \mathbf{g_{k-1}} \end{bmatrix}$$

The matrix **G** is called the generator matrix for C, because the rows of G generate the (n, k) linear code C. Let us arrange these k linearly independent code words as the rows of a $k \times n$ matrix as follows:

$$G = \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ g_3 \\ \vdots \\ g_{k-1} \end{bmatrix} = \begin{bmatrix} g_{00} & g_{01} & \cdots\cdots\cdots g_{0,n-1} \\ g_{10} & g_{11} & \cdots\cdots\cdots g_{1,n-1} \\ . & . & \cdots\cdots\cdots\cdots \\ . & . & \cdots\cdots\cdots\cdots \\ . & . & \cdots\cdots\cdots\cdots \\ g_{k-1,0} & g_{k-1,1} & \cdots\cdots\cdots g_{k-1,n-1} \end{bmatrix} \tag{6.49}$$

Example 6.8 Generation of code word.

$$G = \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ g_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Let $u = [u_0, u_1, u_2, u_4] = [1\ 0\ \ 0\ 1]$ is the message to be encoded, its corresponding codeword would be according to $\mathbf{v} = u_0\mathbf{g_0} + u_1\mathbf{g_1} + \ldots\ldots\ldots + u_{k-1}\,\mathbf{g_{k-1}} = \mathbf{u.G}$,

$$\mathbf{V} = \mathbf{1.g_0 + 0.g_1 + 0.g_2 + 1.g_3} = \mathbf{(1\ 1\ 0\ 1\ 0\ 0\ 0) + (\ 1\ 0\ 1\ 0\ 0\ 0\ 1)}$$
$$= \mathbf{(\ 0\ \ 1\ \ 1\ 1\ \ 0\ \ 0\ \ 1)}$$

Solution A desirable property of LBC is the systematic property in which the parity bits are added to the end of the message bits. For an (n, k) code, the first k bits are identical to the message bits, and the remaining $(n - k)$ bits of each code word are linear combinations of the k information bits. A LBC with this structure is referred to as a linear systematic block code.

A linear systematic (n, k) code is completely specified by a $(k \times n)$ matrix G of the following form:

$$\mathbf{G} = \begin{bmatrix} \mathbf{g_0} \\ \mathbf{g_1} \\ \mathbf{g_3} \\ \vdots \\ \mathbf{g_{k-2}} \\ \mathbf{g_{k-1}} \end{bmatrix} = \left[\begin{array}{c} \overbrace{\begin{matrix} p_{00}\,p_{01}\,p_{02}\cdots\cdots p_{0,\,n-k-1} \\ p_{10}\,p_{11}\,p_{12}\cdots\cdots p_{1,\,n-k-1} \\ ; \\ ; \\ ; \\ ; \\ ; \\ p_{k-1,0}\,p_{k-1,1}\cdots\cdots p_{k-1,\,n-k-1} \end{matrix}}^{\text{parity matrix k xn-k}} \\ \text{p}_{\text{ij}} = 0 \text{ or } 1 \end{array}\right| \overbrace{\begin{matrix} 1\ 0\ 0\cdots\cdots 0 \\ 0\ 1\ 0\cdots\cdots 0 \\ \vdots \\ 0\ 0\ 0\cdots\cdots 1 \end{matrix}}^{\text{kxk identity matrix}} \Bigg| \qquad (6.50)$$

The codeword output for a systematic encoder is of the form, $C_1 = U_1 .G = U_1 . [\ P|\ I_k]$

$$= [u_{i0,}\, u_{i1,}\, \ldots u_{i(k-1),} p_0, p_1, \ldots p_{n-k-1}],\ \text{and}$$
$$p_j = u_{i0}p_{0j} + \ldots + u_{i(k-1)}\,p_{(k-1)j}\,, j = 0,1,\ \ldots, n-k-1 \qquad (6.51)$$

Given the generator matrix $G = \left[\begin{array}{ccccccc} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right]$ and

$\mathbf{u} = (u_0, u_1, u_2, u_3)$ be the message to be encoded.

Then find the corresponding code word.

Let $v = (\mathbf{v_0, v_1, v_2, v_3, v_4, v_5, v_6})$ be the code word to be determined.

We know, $\mathbf{v = u.G} = (u_0, u_1, u_2, u_3)$.

By matrix multiplication, $= \left[\begin{array}{ccccccc} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right]$

We obtain, $v_6 = u_3$, $v_5 = u_2$, $v_4 = u_1$, $v_3 = u_0$, $v_2 = u_1 + u_2 + u_3$, $v_1 = u_0 + u_1 + u_2$, and

$$v_0 = u_0 + u_2 + u_3$$

Now, consider the message is (1 1 1 0), then the corresponding codeword

$$v = (0\ \ 1\ 0\ 1\ 1\ 1\ 0),$$

Table 6.1 shows the generated linear block codes with $k = 4$ and $n = 7$.

Table 6.1 *Linear block codes with k = 4, n = 7*

Message	Codeword	Message	Codeword
0000	0000000	0001	1010001
1000	1101000	1001	0111001
0100	0110100	0101	1100101
1100	1011100	1101	0001101
0010	1110010	0011	0100011
1010	0011010	1011	1001011
0110	1000110	0111	0010111
1110	0101110	1111	1111111

The linear block codes are typically implemented using n–k modulo–2 adders tied to the appropriate stages of the shift register. The resultant parity bits generated are appended at the end of the data bits (message bits) to form the code word. For an example, we can show the modulo–2 implementation for generating (7,4) binary code with the generator matrix,

$$G = \left[\begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{array}\right]$$

Using Eq. (6.51), $p_0 = u_{i0}\, p_{01} + u_{i1}\, p_{11} + u_{i2}\, p_{21} + u_{i3}\, p_{31} = u_{i0} + u_{i1}$, Similarly. $p_1 = u_{i0} + u_{i3}$, and $p_2 = u_{i1} + u_{i2}$. The implementation using shift register to generate the parity bits is shown in Fig. 6.29 and the code word is simply the combination of message bits appended with the parity bits at the end as $[u_{i0}\, u_{i0}\, u_{i0}\, u_{i0}\, p_0 p_1\, p_2]$. The switch position is in downward direction for the output of the systematic bits $u_{ij}, j = 0, 1, 2,3$ of the code and is in the upward position to output the parity bits $p_j, j = 0,1,2$ of the code.

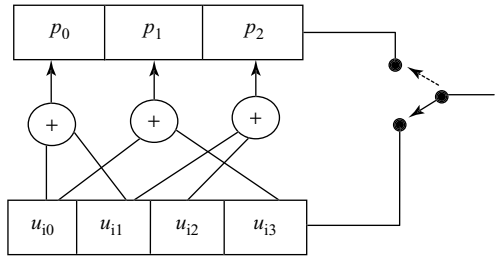


Fig. 6.29 *Implementation of (7,4) binary code word*

6.7.3 Parity Check Matrix and Syndrome Testing

For any (k x n) matrix **G** with k linearly independent rows, there exists an (n–k) x k matrix **H** with (n–k) linearly independent rows such that any vector in the row space of **G** is orthogonal to the rows of H and any vector that is orthogonal to the rows of **H** is in the row space of **G**. An n-tuple **v** is a code word in the code generated by **G** if and only if **v**. $\mathbf{H_T} = 0$, this **H** matrix is called a **parity-check matrix** of the code.
The parity check matrix is used to decode linear block codes with generator **G**. The parity check matrix **H** corresponding to a generator matrix $\mathbf{G} = [\ \mathbf{I_k} \mid \mathbf{P}]$ is defined as

$$H = [P^T \mid I_{n-k}]$$

$$H = \left[\begin{array}{c|c} p_{00}p_{10}p_{20}\cdots\cdots p_{k-1,0} & 1\,0\,0\cdots\cdots 0 \\ p_{01}p_{11}p_{21}\cdots\cdots p_{k-1,1} & 0\,1\,0\cdots\cdots 0 \\ \vdots & \vdots \\ p_{0,n-k-1}p_{1,n-k-1}\cdots\cdots p_{k-1,n-k-1} & 0\,0\,0\cdots\cdots 1 \end{array}\right]$$

Let h_j be the j^{th} row of **H**, so that, $\mathbf{g}_i \cdot \mathbf{h}_j = p_{ij} + p_{ij} = 0$, for $0 \le i < k$ and $0 \le j < n - k$

It implies that, $$\mathbf{G.H^T = 0} \tag{6.53}$$

If R be the received signal resulting from the transmission of codeword V, in the absence of any channel error, $\mathbf{R\,[r_0, r_1, \ldots.r_{n-1}) = V[\,v_0, v_{1,} \ldots v_{n-1}]}$, otherwise **R= V+ e**, where e is the error vector = $[e_0, e_1, \ldots.., e_{n-1})$ indicated the bit of error during transmission. Let us define the syndrome of *R* as,

$$\mathbf{S = R\,H^T = (s_0, s_1, \ldots..,s_{n-k-1})} \tag{6.54}$$

1. **s** = 0, if and only if **R** is a code word and receiver accepts R as the transmitted code word.
2. s ≠ **0** if and only if **R** is not a code word and the presence of errors has been detected
3. When R contains error, but S = $\mathbf{R.H}^T$ = 0, error patterns of this kind is undetectable. As there are $2^k - 1$ non-zero codewords, there are $2^k - 1$ undetectable error patterns

Now using, H= $[P^T \mid I_{n-k}]$ and S= $R.H^T$

$$\begin{aligned} s_0 &= r_0 + r_{n-k}\,p_{00} + r_{n-k+1}\,p_{10} + \ldots\ldots\ldots + r_{n-1}\,p_{k-1,0} \\ s_1 &= r_1 + r_{n-k}\,p_{01} + r_{n-k+1}\,p_{11} + \ldots\ldots\ldots + r_{n-1}\,p_{k-1,1} \\ &\vdots \\ &\vdots \\ s_{n-k-1} &= r_{n-k-1} + r_{n-k}\,p_{0,n-k-1} + r_{n-k+1}\,p_{1,n-k-1} + \ldots\ldots\ldots + r_{n-1}\,p_{k-1,n-k-1} \end{aligned} \tag{6.55}$$

The circuit diagram for generating the syndrome is shown in Fig. 6.30.

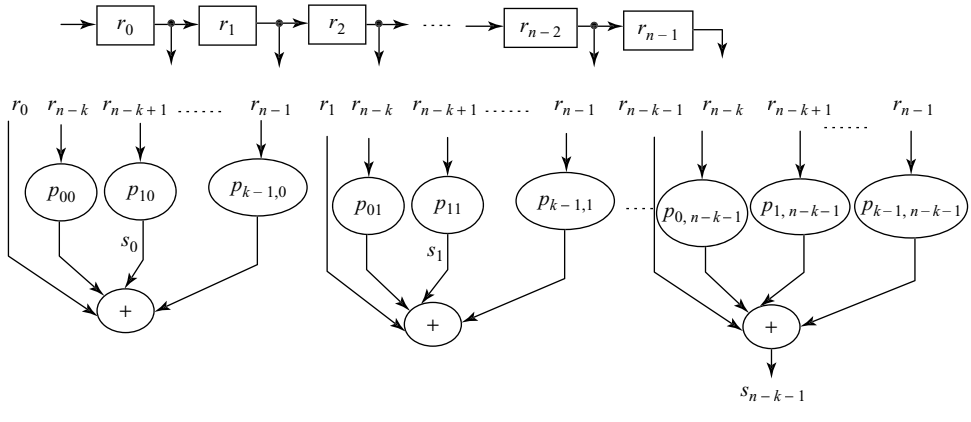


Fig. 6.30 *Syndrome circuit for a linear systematic (n,k) code*

The syndrome S is a function of error only but not the transmitted code word V, as

$$S = RH^T = (V + e)\, H^T = V.H^T + e.H^T = 0 + eH^T = e.H^T \tag{6.56}$$

which corresponds to $n-k$ equations in n unknowns. There are 2^k possible error patterns that can produce a given syndrome. In general the probability of error is small and independent for each bit. So the most likely error pattern considered for minimal weight corresponding to the least number of error within the channel.

Hamming Weight Hamming weight of a codeword v is denoted by $w(v)$, which is defined as the number of nonzero components of v.

For example, the Hamming weight of $v = (1\ 0\ 0\ 0\ 0\ 1\ 1)$ is 3.

Hamming distance of two codeword v and w as defined earlier is defined by d(v, w), is the number of places they differ.

Given the block code C, the minimum distance of C is $d_{\min}$, defined as

$$d_{\min} = \min \{\, d(v,w) : v, w \in C, \text{ and } v \neq w \,\} \tag{6.57}$$

As the Hamming distance between two n-tuples, **v** and **w**, is equal to the Hamming weight of the sum of **v** and **w**, that is

$$d(\mathbf{v}, \mathbf{w}) = w(\mathbf{v} + \mathbf{w}) \tag{6.58}$$

Then the Hamming distance between two code vectors in C is equal to the Hamming weight of a third vector in C

$$\begin{aligned} d_{\min} &= \min \{\, w(v,w) : v, w \in C, \text{ and } v \neq w \,\} \\ &= \min \{\, w(y) : y \in C, y \neq 0 \,\} = w_{\min} \end{aligned} \tag{6.59}$$

$w_{\min}$ is called the minimum weight of the linear code C. The random error detecting capability of a block code with minimum distance $d_{\min}$ is $d_{\min} - 1$ and the error correcting capability is less than $\frac{\lfloor d_{\min} - 1 \rfloor}{2}$.

6.7.4 Cyclic Codes

Cyclic codes are a subset of the linear block codes, which satisfy the cyclic shift property. If $C = [c_{n-1}, c_{n-2}, \ldots\ldots, c_1, c_0]$ is a valid code word has gone through one cyclic shift, then $C = [c_{n-2}, c_{n-3}, \ldots, c_0, c_{n-1}]$, is also a code word. An (n,k) linear code **C** is cyclic if every cyclic shift of a codeword in **C** is also a codeword in **C**. That is all cyclic shifts of **C** are code words.

If $c_0\ \ c_1\ \ c_2\ \ \ldots c_{n-2}\ \ c_{n-1}$ is codeword, then

$c_{n-1}\ \ c_0\ \ c_1\ \ \ldots c_{n-3}\ \ c_{n-2}$

$c_{n-2}\ \ c_{n-1}\ \ c_0\ \ \ldots c_{n-4}\ \ c_{n-3}$

$\vdots\ \ \ \vdots\ \ \ \vdots\ \ \ \ \vdots\ \ \ \ \vdots$

$c_1\ \ c_2\ \ c_3\ \ \ldots c_{n-1}\ \ c_0$ are all code words.

The (7,4) linear block codes given in Table 6.1 are cyclic.

Cyclic codes are generated by Generator Polynomial rather than generator matrix. The generator polynomial G(x) for an (n,k) cyclic code has degree $(n-k)$ and is represented as,

$$G(x) = g_0 + g_1 x + g_2 x^2 \ \ldots\ldots + g_{n-k} x^{n-k} \tag{6.60}$$

where $g_0 = g_{n-k} = 1$ and $g_i = 0$ or 1, and the power of x corresponds to the bit position.

The k bit message sequence $(u_0, u_1, u_{2, \ldots\ldots}, u_{k-1})$ is also represented by a polynomial $U(x)$ and is represented as,

$$U(x) = u_0 + u_1 x + u_2 x + \ldots\ldots + u_{k-1} x^{k-1} \tag{6.61}$$

The codeword $C(x) = G(x)\ U(x) = c_0 + c_1 + c_2 + \ldots\ldots + c_{n-1}\ x^{n-1}$ (6.62)

The codeword described by a polynomial $C(x)$ is a valid cyclic code if and only if $G(x)$ divides $C(x)$ with no remainder, i.e., $C(x)/\ G(x) = Q(x)$

where, $Q(x)$ is a polynomial of degree less than k.

Example,

1010011 $1 + x^2 + x^5 + x^6$, 0101110 $x + x^3 + x^4 + x^5$

Within a systematic block codes of (n, k), the first k bits represent the message bits and the remaining $n–k$ bits are the parity bits. The cyclic code is represented in a systematic form by multiplying the message polynomial $U(x)$ with x^{n-k} which yields,

$$U(x)\ x^{n-k} = u_0\ x^{n-k} + u_1\ x^{\ n-k-1} + \ldots\ldots\ldots + u_{k-1}\ x^{n-1} \tag{6.63}$$

It shifts the message bits to the k rightmost position of the code word polynomial. Next dividing Eq. (6.63) by $G(x)$,

$$\frac{U(x)\ x^{n-k}}{G(x)} = Q(x) + P(x)/G(x)$$

where $Q(x)$ is a polynomial of degree (k–1) and $P(x)$ is a polynomial of degree ($n–k$–1), and can be expressed as, $P(x) = p_0 + p_1\ x + \ldots\ldots\ldots + p_{n-k-1}\ x^{n-k-1}$ (6.64)

Thus, $U(x)\ x^{n-k}\ = Q(x)\ G(x) + P(x)$ (6.65)

So, adding $P(x)$ to both sides, $U(x)\ x^{n-k} + P(x) = Q(x)\ G(x)$ (6.66)

$U(x)\ x^{n-k} + P(x)$ is a valid code word when dividing by $G(x)$ will left no remainder.

Combining Eqs. (6.65) and (6.66),

$$U(x)\ x^{n-k} + P(x) = p_0 + p_1\ x + \ldots + p_{n-k-1}\ x^{n-k-1} + u_0\ x^{n-k} + u_1 x^{\ n-k-1} + \ldots + u_{k-1}\ x^{n-1} \tag{6.67}$$

So the generation of systematic code word polynomial is a three steps process:

1. Multiplying the message polynomial $U(x)$ with x^{n-k} to get $U(x)\ x^{n-k}$.
2. Dividing $U(x)\ x^{n-k}$ by $G(x)$ to get $P(x)$ and $Q(x)$.
3. Adding $P(x)$ to $U(x)\ x^{n-k}$.

If $e(x)$ is the error polynomial of degree (n–1) with error coefficients as 1 at the place of occurrence of the error, then the received code word for cyclic code word is written as,

$$R(x) = C(x) + e(x) = U(x)\ G(x) + e(x) \tag{6.68}$$

When $R(x)$ will be divided by $G(x)$, a syndrome polynomial $S(x)$ of degree ($n–k$–1) will be generated so that,

$$e(x) = G(x)\ S(x) \tag{6.69}$$

Therefore, the syndrome polynomial $S(x)$ is equivalent to the error polynomial $e(x)$ modulo $G(x)$ and can be easily implemented using feedback shift register, resulting in a low cost solution.

References

[1] Qureshi, Shahid U.H. , *Adaptive Equalization*, Proceedings of the IEEE, *Vol.* 73. No. 9, September 1985, pp. 1349–1387.

[2] Dong, X., and N.C. Beaulieu, *Optimal Maximal Ratio Combining with Correlated Diversity Branches*, IEEE Communications Letter, vol. 6, no. 1, pp. 22–24, 2002.

[3] Brennan, D.G., *Linear Diversity Combining Techniques*, Proc. IRE, vol. 47, pp. 1075–1102, 1959.

[4] Shah, A., and A.M. Haimovich, *Performance Analysis of Maximal Ratio Combining and Comparison with Optimum Combining for Mobile Radio Communications with Co-channel Interference*, *IEEE Transactions on Vehicular Technology,* Vol. 49, no. 4, July 2000, pp. 1454–1463.

[5] Proakis, J.G., *Digital Communication*, 4th edition, New York, 2001.
[6] Winters, J.H., *Optimum Combining in Digital Mobile Radio with Cochannel Interference*, *IEEE Trans. Veh. Technol.*, vol. VT-33, pp. 144–155, Aug. 1984.
[7] Kam, P.Y., *Bit Error Probabilities of MDPSK Over the Non-selective Rayleigh Fading Channel with Diversity Reception*, *IEEE Trans. Commun.*, Vol. 39, pp. 220–224, Feb. 1991.
[8] Beaulieu, N.C., and A. A. Abu-Dayya, "*Analysis of Equal Gain Diversity on Nakagami Fading Channels*," *IEEE Trans. Commun.*, Vol. 39, pp. 225–233, Feb. 1991.
[9] Cioffi, J., and T. Kailath, *Fast, Recursive-Least-Square Transversal Filter for Adaptive Filtering*, IEEE Trans. Signl. Proc., Vol 32, No.2 , pp. 304–337, April 1984.
[10] Salz, J., *Optimum Mean Square Decision Feedback Equalization*, Bell Syst. Tech., Journal, Vol. 52, pp. 1341–1373, Oct 1973.
[11] Cioffi, J.M., G.P. Dudevour, V. Eyuboglu, and G.D. Formey, Jr., *MMSE Decision-Feedback Equalizers and Coding*, *Part I: Equalization Results*, IEEE Trans. Communication, pp. 2582–2594, Oct. 1995.
[12] Cioffi, J.M., G.P. Dudevour, V. Eyuboglu, and G.D. Formey, Jr., *MMSE Decision-Feedback Equalizers and Coding, Part II: Equalization Results*, IEEE Trans. Communication, pp. 2595–2604, Oct. 1995.
[13] Mark, J.W. and Zhuang W., *Wireless Communication and Networks*, PHI, New Delhi 2005.
[14] Haykin, S., *Adaptive Filter Theory*, Prentice Hall, Inc., 2001.
[15] Dong, X., and N. C. Beaulieu, *Optimal Maximal Ratio Combining with Correlated Diversity Branches*, *IEEE Communications Letter*, Vol. 6, no. 1, pp. 22–24, 2002.

Questions for Self-Test

6.1 Describe the basic concept of equalization process to combat ISI effect in a fading channel.

6.2 What are the different types of equalization? Mention in which situation each of the equalizers would be useful.

6.3 Show that equalizer process enhances the noise.

6.4 Let the complex impulse response of the equalizer is $h_{eqz}(t) = \Sigma_n c_n \delta(t - nT)$, c_n is the complex filter coefficients of the equalizer. Show that in absence of noise, $F^*(-f)\, H_{eqz}(f) = 1$, where $F^*(-f)$ and $H_{eqz}(f)$ are the Fourier transform of the function $f^*(t)$ and $h_{eqz}(t)$ respectively.

6.5 Consider a linear transversal equalizer with $(2M+1)$ taps and tap coefficients w_{-M} to w_M, if $D(z)$ is the z-domain transfer function of the equalizer and $R(z)$ is the input of the equalizer coming from effective channel h_{eff}, then find the output of the equalizer $Y(z)$.

6.6 What is zero forcing equalizer? Explain the operation of ZF equalizer with effective channel response c_n and equalizer response d_n.

6.7 Consider the discrete frequency channel response in a multipath wireless propagation is $h(n) = [.1, .9, .3]$. Design the 3-tap zero forcing equalizer for the system for ISI free transmission. Plot the frequency domain channel and equalization response.

6.8 Repeat Problem 6.6 for 5 tap and 7 tap ZF equalizers. Also plot the frequency response curves. Compare the results obtained in problems 6.6 and 6.7. Comment on the results.

6.9 Find the bit error probabilities for 3-, 5-, and 7- tap ZF equalizers. Consider the effective channel impulse response $h(n) = [1, .5] = \delta(n) + 0.5\, \delta(n-1)$.

6.10 Establish the relationship of Wiener-Hopf equation for MMSE equalizer.

6.11 Find the equalizer coefficients using 3-tap delay channel with response $h(n) = [h_{-1} = 0.1, h_0 = 0.9$ and $h_1 = 0.3]$ and $h(n) = [0\ 1.0\ .5]$ using minimum mean square error (MMSE) algorithm. Compare the results of the first part with respect to Problem 6. 6.

6.12 What is adaptive equalizer? Describe the basic principle of adaptive ness in respect of MMSE equaizer.

6.13 Plot the frequency response for MMSE equalizer using 3-, 5- and 7- tap channel with response $h(n) = [0\ 1.0\ .5]$ and $h(n) = [.1\ .9\ .3]$.

6.14 What are the different types of diversity techniques generally used? Describe the basic principle of each type.

6.15 What are the different types of space diversity?

6.16 Describe the basic principle for selection diversity technique.

6.17 Show that the maximum received SNR = $\gamma_{Max} = A^2 E_b / N_0 = \text{Max}(\gamma_1, \gamma_2, \ldots\ldots\gamma_{M-1})$, where E_b/N_0 is the transmitted bit energy-to-noise ratio of the channel for the selection diversity. Considering uncorrelated branches of diversity channel and identically distributed (iid) fading, find the pdf of the SNR in the diversity reception for Rayleigh fading channel.

6.18 Plot the pdf with respect to SNR in selection diversity for Rayleigh fading for different branch. Comment on the result.

6.19 Repeat Problem of 6.15 for Maximal ratio combining (MCR) and plot the pdf for this. Compare the plot of pdf obtained for SC and MCR.

6.20 Show that the weights w_i that maximize SNR(γ_{max}) under MRC are $A_i^2 = r_i^2$ /N for N the common noise power on each branch. Also show that with these weights $\gamma_{max} = \sum_i \gamma_i$.

6.21 Derive the average probability of bit error for BPSK and QPSK modulations with iid Rayleigh fading on each branch of MRC diversity. Derive the same for EGC technique.

6.22 Compare the average probability of bit error for BPSK modulation under no diversity, two branch SC diversity, two branch MRC diversity and two branch EGC diversity assuming iid Rayleigh fading on each branch. The SNR for first channel is 10 dB and second channel is 20 dB. Compare the results and comment.

6.23 Describe the technique to combat multipath effect in DS-CDMA system using RAKE receiver.

6.24 How does channel coding help in receiving the signal over noisy channel?

6.25 If $\mathbf{u} = (u_0, u_1, \ldots, u_{k-1})$ is the message to be encoded, the corresponding code word can be given as follows:

$$\mathbf{v} = u_0\mathbf{g_0} + u_1\mathbf{g_1} + \ldots\ldots\ldots + u_{k-1}\,\mathbf{g_{k-1}} = \mathbf{u.G.}\ \text{If}$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{g_0} \\ \mathbf{g_1} \\ \mathbf{g_2} \\ \mathbf{g_3} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1\,0 & 0 & 0 \\ 0 & 1 & 1 & 0\,1 & 0 & 0 \\ 1 & 1 & 1 & 0\,0 & 1 & 0 \\ 1 & 0 & 1 & 0\,0 & 0 & 1 \end{bmatrix},$$

, find the generated codeword.

6.26 A linear systematic (n, k) code is completely specified by a $(k \times n)$ matrix G of the following form,

$$\mathbf{G} = \begin{bmatrix} \mathbf{g_0} \\ \mathbf{g_1} \\ \mathbf{g_2} \\ \vdots \\ \mathbf{g_{k-2}} \\ \mathbf{g_{k-1}} \end{bmatrix} = \left[\begin{array}{l} p_{00}\,p_{01}\,p_{02}\cdots\cdots p_{0,\,n-k-1} \\ p_{10}\,p_{11}\,p_{12}\cdots\cdots p_{1,\,n-k-1} \\ ; \\ ; \\ ; \\ ; \\ ; \\ p_{k-1,0}\,p_{k-1,1}\cdots\cdots p_{k-1,\,n-k-1} \end{array}\right.\left|\begin{array}{l} \text{parity matrix k xn-k} \\ 1\,0\,0\cdots\cdots 0 \\ 0\,1\,0\cdots\cdots 0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 0\,0\,0\cdots\cdots 1 \end{array}\right.\begin{array}{l} \text{kxk identity matrix} \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{array}$$

$p_{ij} = 0$ or 1

Consider a (3,1) linear block code where each codeword consists of 3 data bits and one parity bit. Find (a) all codewords in this code (b) the minimum distance of the code.

6.27 Considering a (7,4) code with generator matrix

$$G = \begin{bmatrix} 0 & 1 & 0 & 1\,1 & 0 & 0 \\ 1 & 0 & 1 & 0\,1 & 0 & 0 \\ 0 & 1 & 1 & 0\,0 & 1 & 0 \\ 1 & 1 & 0 & 0\,0 & 0 & 1 \end{bmatrix}$$

Find the all codewords of the code.

(a) What is the minimum distance of the code?
(b) Find the parity check matrix of the code.
(c) Find the syndrome for the received vector R = [1 1 0 1 1 1]
(d) Assuming an information bit sequence of all 0s, find all the minimum weight error patterns e that result in a valid codeword that is not the all zero codeword.

6.28 If all the Hamming codes have a minimum distance of 3, what is the error correction and error detection capability of a Hamming code?

6.29 The (7,4) cyclic hamming code has a generator polynomial $g(x) = 1 + x^2 + x^3$,

(a) Find the generator matrix of this code in systematic form
(b) Find the parity check matrix for the code
(c) Suppose the codeword is C= [110 0011] is transmitted through the noisy channel, and the received codeword is $C^{'}$ = [11 010011], find the syndrome polynomial associated with this received codeword.

6.30 The parity check matrix is used to decode linear block codes with generator **G**. The parity check matrix **H** corresponding to a generator matrix $\mathbf{G = [I_k \mid P]}$ is defined as

$$H = [P^T \mid I_{n-k}]$$

If **R** be the received signal resulting from the transmission of codeword **V**, in the absence of any channel error, $\mathbf{R\ [r_0, r_1, \ldots r_{n-1}) = V[\ v_0, v_1, \ldots v_{n-1}]}$, otherwise $\mathbf{R = V + e}$, where e is the error vector = $[e_0, e_1, \ldots, e_{n-1})$ indicated the bit of error during transmission. Let us define the syndrome of R as $\mathbf{S = R\ H^T = (s_0, s_1, \ldots, s_{n-k-1})}$.

Draw the circuit diagram for generating the syndrome.

Multiple Access Techniques in Wireless Communications

7

Introduction

Sharing of resources among multiple users in any communication system (wireless or wired) is referred as multiple access technology. In wireless communication, resource means radio frequency channel. The objective of multiple access technology is to provide this radio resource simultaneously to the multiple users maintaining the acceptable limit of interference for maximum utilization of the resources. The wireless cellular communication systems pass through several generations, from first to third. The first generation cellular systems use frequency division multiple access (FDMA), the second generation mainly use time division multiple access (TDMA) and the third generation cellular systems use the code division multiple access (CDMA). For each development process, the main focus was to enhance the system capacity and support of higher data rate. Again, the cellular industry has begun to explore the conversion of analog systems to digital systems from first generation to second generation since the late 1980s. In the year 1989, the cellular telecommunication industry association (CTIA) adopted the TDMA technology over narrowband FDMA technology. Meanwhile with the growing competition of technologies, Qualcomm introduced the CDMA technology for the carrier selection. Multiple access strategies based on the orthogonal principle among the competing transmissions are collision free. Orthogonality can be in the form of frequency, time and code division multiple access. The two major systems are the TDMA and CDMA techniques to use the radio resources. Space Division Multiple Access (SDMA) is another technology used in wireless communication systems. In this chapter, we shall discuss about the principles of operation of the different access technologies like FDMA, SDMA, TDMA and CDMA with their pros and cons.

7.1 FREQUENCY DIVISION MULTIPLE ACCESS TECHNOLOGY (FDMA)

In frequency division multiple access, the frequency band is divided into slots. Each user gets one frequency slot that is assigned during communication. It could be compared to an AM or FM broadcasting radio where each station has a frequency assignment. For any wireless communication, duplexing is required for simultaneous two-way communication. Frequency division duplexing (FDD) provides two separate bands of frequencies for every user. The band from the base station to the mobile is called the *forward band* (channel), and that from the mobile to the base station is called *reverse band* (channel). A duplexer is used for each user handset and base station to allow simultaneous bi-directional communication, thus increasing the cost of the subscriber unit and base stations. The frequency separation between the forward and reverse channel is fixed for a communication system. To avoid interference between the uplink and downlink channels, the frequency allocations have to be separated by a sufficient amount that can be achieved using two antennas at different frequencies.

Although FDMA implementation is simple, it is wasteful for resources as each user is assigned a channel for communication and remains unused during the non-conversation period. Mainly used for voice communication, FDMA demands good filtering to minimize adjacent channel interference. It is generally implemented in narrowband

systems, not suitable for multimedia communications with various transmission rates. In the narrowband multiple access system, the entire bandwidth is divided into several narrowband channels. The bandwidth of the single channel is narrow with respect to the coherence bandwidth of the channel. Another main problem in the FDMA system is the effect of non-linearity that arises when the power amplifier operates at or near saturation for maximum efficiency. Non-linearity generates the inter-modulation frequencies resulting in inter-modulation distortion, and spreading of frequency is a common phenomenon which again causes adjacent channel interference.

In the FDMA system, a channel is identified by frequency. Figure 7.1 shows the FDMA scheme.

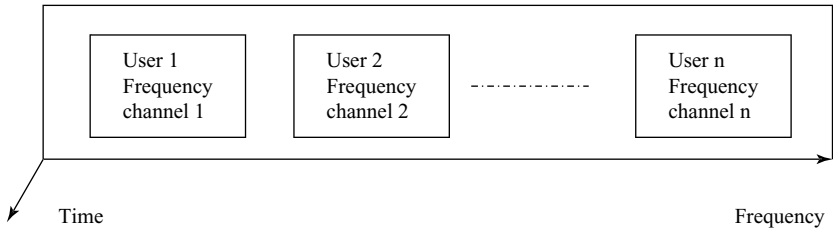


Fig. 7.1 *FDMA scheme*

The first analog cellular system AMPS (Advanced Mobile Phone Systems) in America used FDMA/FDD using a 25 MHz band in each uplink, from 824 to 849 MHz; and a 25 MHz band in each downlink, from 869 to 894 MHz (Fig. 7.2). AMPS uses a channel spacing of 30 KHz with a total capacity of 832 channels. A single user occupies a single channel (two simplex channels) with frequency duplexing using 45 MHz frequency separation.

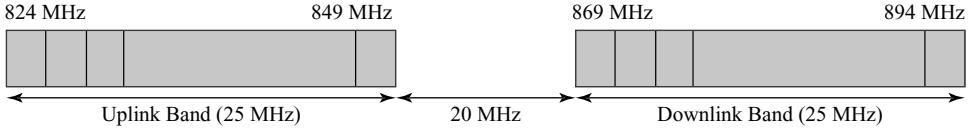


Fig. 7.2 *Frequency allocation in AMPS*

7.2 TIME DIVISION MULTIPLE ACCESS (TDMA)

The time division multiple access is divided into two methods—wideband TDMA (WTDMA), where the entire frequency band (BW_t) is allocated to users into different time slots, and narrowband TDMA, where the frequency spectrum band is partitioned in narrow frequencies, which in turn are divided into a number of time slots as shown in Fig. 7.3. Users are allowed to use it only in predefined intervals of time for each frame. Thus, TDMA demands synchronization among the users.

With TDMA, the receiver must be able to synchronize the received signal within a slot time. A convenient way to extract timing information is to use matched filtering or co-relation detection. To provide required separation between the uplink and downlink channels, TDMA can use TDD (Time Division Duplex) or FDD. Instead of frequency, TDD uses time for both the forward and reverse links. Half of the time slot is used for the forward link, and the other half is used for the reverse link. Many users in time domain use a single radio channel with an assigned time slot to each user. Each duplex channel has time slots in the forward and reverse time slots to facilitate both ways of communication. This time separation is small such that users perceive simultaneous transmission and reception. In TDD, a duplexer is not required, and so a subscriber unit is less costly.

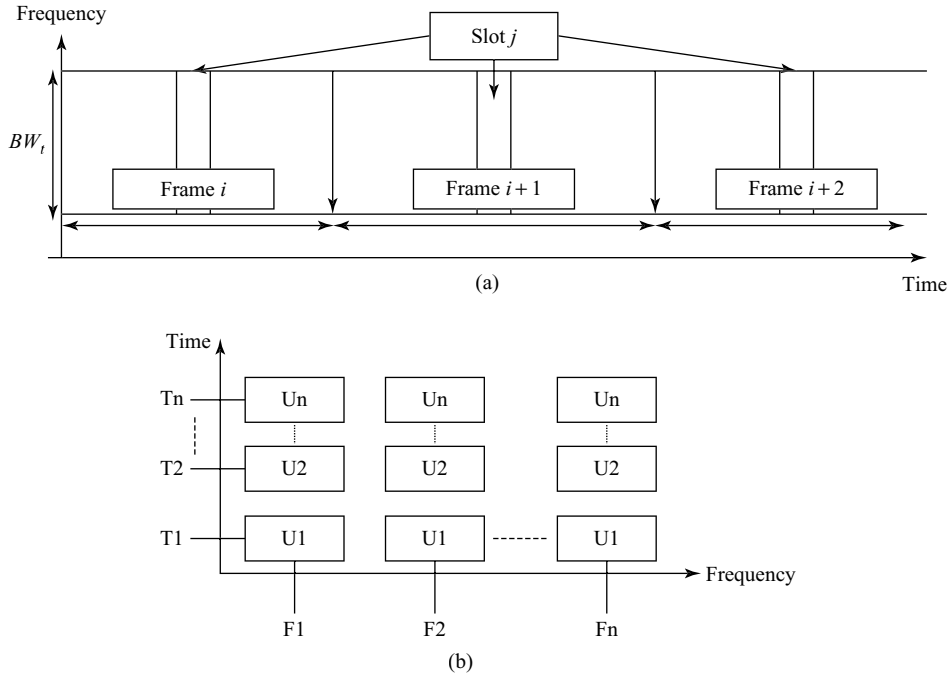


Fig. 7.3 *(a) Wideband TDMA, (b) Narrowband TDMA*

In TDMA, a channel means a particular time slot that is cyclically repeating to every user. A TDMA frame consists of many time slots. The transmission from different users is interlaced into a repeating frame structure as shown in Fig. 7.4.

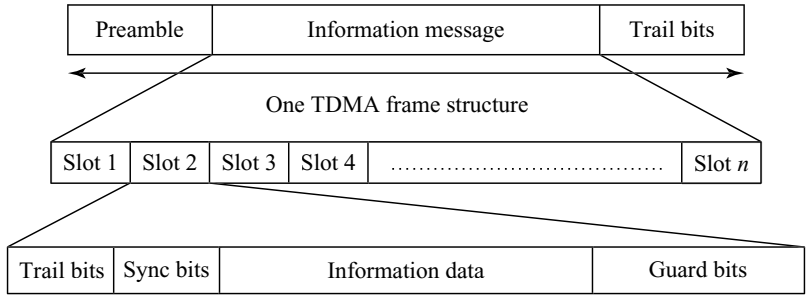


Fig. 7.4 *TDMA frame structure*

The preamble contains the address and synchronization information for identifying the base station and subscriber stations. For transmitting information message, users are allowed to assign a time slot. This slot assignment can be fixed or dynamic. For a fixed type of slot assignment, users need to synchronize their respective slot assignments and this is known as *synchronous TDMA (STDMA).* In case of dynamic assignment, slot allocation to a user is dynamic during the packet transmission and this mode is referred as *asynchronous TDMA (ATDMA).* Unlike STDMA, the frame length in ATDMA varies depending on the number of active users in the frame. Again, the numbers of time slots depend on the modulation technique and available bandwidth. Guard times are used in a TDMA frame to allow synchronization of the receivers

between different slots and frames. This guard time needs to be optimized in order to avoid adjacent channel interference. Sharp cutting of a transmitted signal at the end of a time slot is not desirable, as it causes spectral spreading to the adjacent bands.

Data transmission in TDMA is not continuous, but data comes in bursts. This is very helpful in battery power consumption, as the user's transmitter may keep a 'turned-off' state during the no transmission period. The transmission rate in a TDMA system is very high compared to that in an FDMA system, and a TDMA system requires adaptive equalisation. The discontinuous data transmission helps for easy handoff by using the idle time slots. The subscriber may listen the other base stations during this idle time in the TDMA frame.

One problem of the TDMA system is the increased overhead compared to FDMA due to the synchronization process and guard slots for separating users. Receivers require synchronization for each data burst.

GSM (Global system for Mobile) networks use a combination of TDMA/FDD. GSM900 operates at an 890–915 MHz for reverse link (uplink—mobile to base station) and 935–960 MHz for forward link (downlink—base station to mobile station). Frequency-division multiplexing divides each of the frequency ranges into 125 channels (25 MHz / 200 kHz) of 200 kHz bandwidths. One of the channels is used as a guard band, leaving 124 channels available for both transmission and reception. Thus, in GSM, a total of 124 duplex communication channels are produced. Channel 1 is transmitted at the 890.1 MHz ± 100 kHz range, Channel 2 is of the 890.3 MHz ± 100 kHz range, and so on. Each frequency channel in GSM provides eight (8) time slots, one for each radio carrier having a duration of 577 μs. So, radio interfaces of carrier waves can be simultaneously transmitted to eight mobile subscribers at the same frequency channel. Therefore, each transmitter uses the same channel, after 577 μs × 8 = 4.615 ms, and so one TDMA frame duration in a GSM system is 4.615 ms long. Figure 7.5 shows the GSM TDMA/FDD channel assignment scheme.

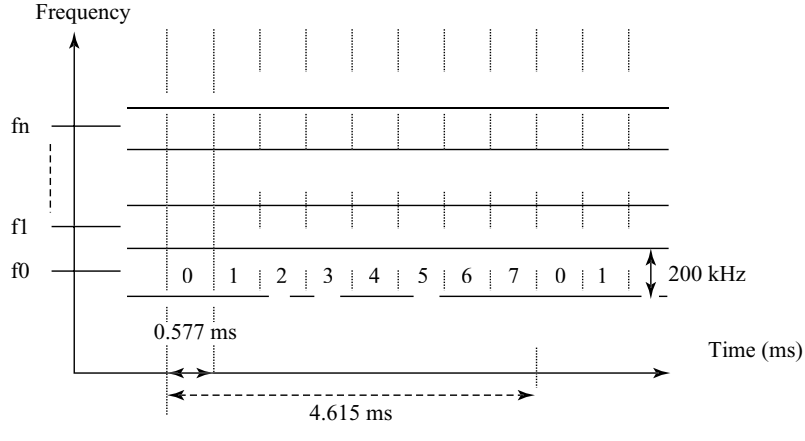


Fig. 7.5 *TDMA/FDD combination in GSM*

7.3 SPACE DIVISION MULTIPLE ACCESS (SDMA)

Space division multiple access controls the radio signal for each user in space. Using different antennas with a directed beam pattern, SDMA serves different users. The ideal application example of this technique is the sectorization of antennas in a cellular environment. In SDMA, a wireless transmitter transmits the modulated signals and accesses the space slot. Another transmitter accesses another space slot such that signals from both transmitters can propagate in two separate spaces without affecting each other. However, there is a limit

of using the number of cells that can be formed to serve mobile users using SDMA as there may arise some interference between cells transmitting signals with the same radio carrier frequency. Figure 7.6 illustrates the SDMA technique.

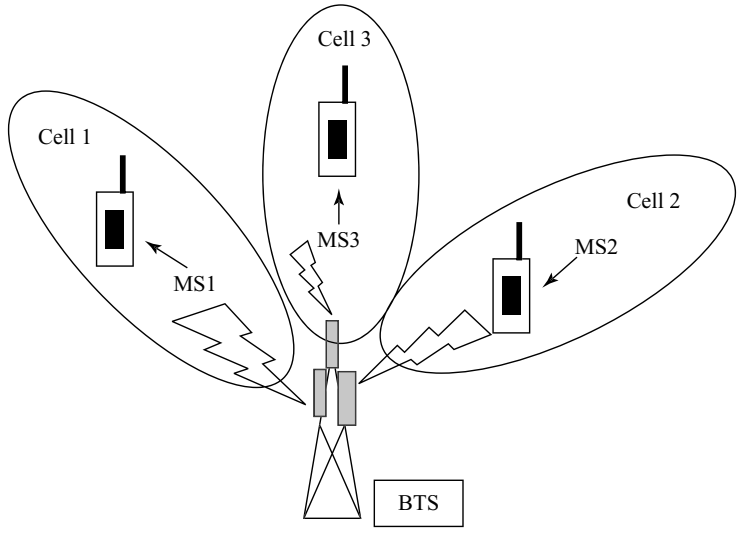


Fig. 7.6 *Three cells in transmission with the same radio frequency channel using SDMA*

SDMA is an alternative technique of increasing the capacity of TDMA/FDMA systems. In this system, frequency reuse remains same while the number of users within a given cell increases by using the directional antenna beam pattern.

7.4 CODE DIVISION MULTIPLE ACCESS (CDMA)

Code division multiple access, is different from the other traditional systems in that it does not allocate frequency or time in user slots but gives the right to use both to all users simultaneously.

To enable this, CDMA uses a technique known as ***spread spectrum***. Each user is assigned a code, which spreads its signal bandwidth in such a way that only the same code can recover it at the receiver end. This method has the property that the unwanted signals with different codes get spread even more by the process, making them appear as a noise to the receiver. Spread spectrum describes any system in which a signal is modulated so that its energy is spread across a frequency range that is greater than that of the original signal.

By spread spectrum, transmission of data occupies a larger bandwidth than necessary. Bandwidth spreading is accomplished before the transmission through the use of a code (PN sequence generated by cyclic code generator) that is independent of the transmitted data. The same code is used to demodulate (de-spread) the data at the receiving end by the PN sequence locally generated. Figure 7.7(a) illustrates the spreading done on the data signal $x(t)$ by the spreading signal $c(t)$ resulting in the message signal $m(t)$ to be transmitted. A CDMA in which the spread spectrum is achieved by directly multiplying the user's baseband signal with a high-rate PN sequence is referred as direct sequence CDMA (DS-CDMA).

Since multiplication in the time domain corresponds to convolution in the frequency domain, a narrowband signal when multiplied by a wideband signal also results in a wideband signal. One way of doing this is to use a binary waveform as a spreading function, at a higher rate than the data signal as shown in Fig. 7.7(b).

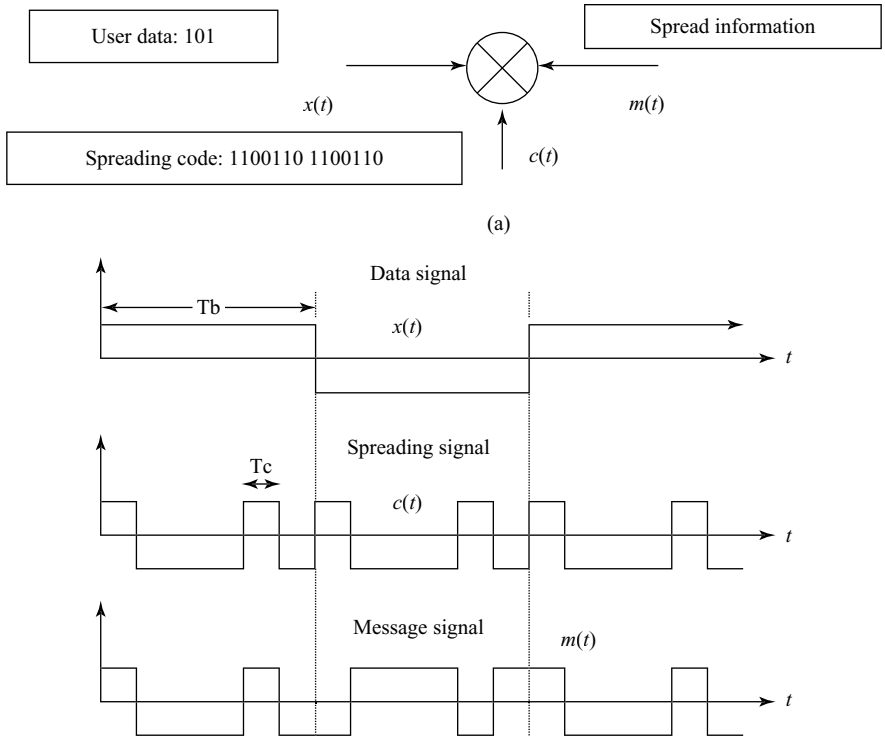


Fig. 7.7 *(a) Basic CDMA technique, (b) Spreading of signal*

In DS-CDMA technology, all the users simultaneously transmit at the same frequency that is different from frequency hopping (FH) spread spectrum where the carrier frequency changes at regular time intervals. These frequencies are selected randomly from a pre-defined group within the available spectrum. Figure 7.8 shows the DS-CDMA trans-receiving technique. Each transmitted signal $s_i(t)$ is modulated by a unique code $c_i(t)$ and is combined at the baseband. The receiver receives the sum of all signals from all communication channels within the RF carrier. The de-spreading is done through multiplication of the received signal by the code associated with the desired channel.

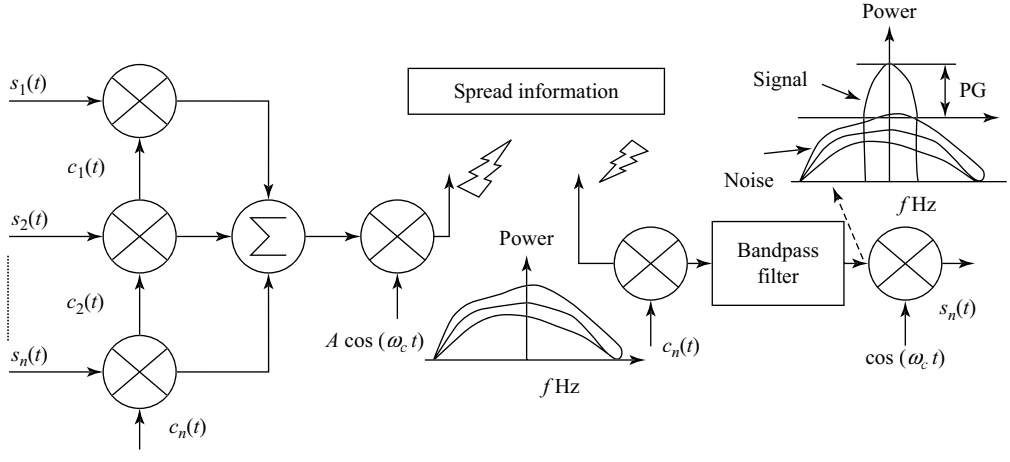


Fig. 7.8 *DS-CDMA trans-receiving technique*

All users in the CDMA system use the same carrier frequency and may transmit at the same time with a different pseudorandom code word. These code words need to be orthogonal to each other. Signal separation is done by the orthogonality property of the codes. For perfect orthogonality between two codes, the cross-correlation factor must be zero for a time separation $\tau = 0$. Theoretically, any CDMA system can support as many users as long as it has unique orthogonal codes. But there is a limit; increasing number of users degrades the system performance due to the increase of noise floor. Spreading codes of different users may not be perfectly orthogonal leading to self-jamming. Orthogonal codes must be perfectly synchronized to achieve signal separation. Another problem of orthogonal codes, is that they do not evenly spread signals across the wide frequency band, but concentrate the signal at certain discrete frequencies.

Power control by the base station in a CDMA system is very important to overcome the effects of multipath propagation and fading, due to the different propagation characteristics transmitting simultaneously. If the power of each mobile station within a cell is not controlled then they do not appear equal to the base station. This effect is known as the **near–far** problem. A user who is physically much close to the base station may have more power than a far-end user. The power control mechanism enables that signals arriving at the receiver have equal power and at a level that satisfies the required signal-to-interference ratio.

If small scale or shadow fading occurs due to the obstruction present within the propagation path of the signal, it will change the received power level. Multiple versions of the signal will be received at the receiver after having constructive and destructive interference as the relative time shifts between the signals. A **RAKE** receiver is used to improve reception quality by collecting time-delayed versions of the required signal–therefore it improves transmission accuracy, especially during soft handoff.

In fast fading, or Rayleigh fading, that occurs due to Doppler shift, the apparent wavelength of the transmitted signal will increase as the mobile moves towards the receiver or it may decrease while the mobile moves away from the receiver. As all the received signals with varied phase shift arrive at the receiver, the signals that are received in phase have peaks of power and reinforce each other. On the other hand, out-of-phase signals may have weak power. This variation of power may be prominent within a very short distance compared to the wavelength of the transmitting signal. That is why power control must be maintained to mitigate the multipath fading effects.

Table 7.1 *Multiple-access techniques in different wireless systems*

Access Techniques	Cellular Wireless systems
FDMA/FDD	Used in 1G analog AMPS (Advanced Mobile Phone system) in North America
TDMA/FDD	Widely used in 2G digital cellular networks in GSM (Global System for Mobile) in Europe, Asia, etc., USDC (US Digital Cellular) in USA, PDC (Pacific Digital Cellular) in Japan
FDMA/TDD	Cordless Telephones CT2, and DECT (Digital European Cordless Telephone) in Europe and Asia
CDMA/FDD	Mainly used in 2G CDMA one or IS-95 (Interim Standard CDMA cellular networks) used in North America, Korea, Japan, etc.
CDMA/FDD CDMA/TDD	3G Cellular Networks UMTS (Universal Mobile Telecommunication System) under the 3GPP (Third Generation Partnership Project) in the evolution path of GSM-GPRS-UMTS, based on Wideband CDMA. cdma 2000 under 3GPP2 developed in America in the evolution path of IS-95-IS-95G-cdma2000

Another important advantage of the CDMA system is the soft handoff. As mobile users operate at the same frequency within a cell, it is possible to communicate with multiple base stations simultaneously. During the soft-handover procedure, a connection is made to the target base station before the connection is broken with the original base station. This is in contrast to the TDMA/FDMA system where the connection is dropped first and then a new connection is made as cells operate at different frequencies. Using macroscopic spatial diversity, MSC can monitor a particular user from more than one base station as CDMA uses co-channel cells.

As CDMA is interference limited—if a user does not transmit, it does not add any interference with other active users, thus leading to high resource utilization via statistical multiplexing for on-off voice traffic and bursty data traffic. For a CDMA system, the minimum cell cluster size may be $N = 1$ and maximum frequency reuse is possible, thus reducing complexity of frequency planning.

CDMA has more flexibility than TDMA in supporting multimedia services, but comes at a high price for high complexity of the transmitter and Rake receiver, than in TDMA and FDMA systems.

7.4.1 Processing Gain in CDMA System

In spread spectrum, the transmitted signal is retrieved at the receiver end with its original bandwidth after de-spreading the signal. Both the PN sequences used at the transmitter and receiver have to be identical and synchronized. The other signal at the receiver input behaves like a noise and remains spread. The energy of these signals is reduced in the narrowband filter after the correlation process. The de-spreading process strengthens the desired signal in relation to other signals. Here lies the significance of defining processing gain.

For direct-sequence spread spectrum, bits of the spreading signal are called chips. On Fig. 7.7b, T_b represents the period of one data bit and T_c represents the period of one chip. The chip rate, $1/T_c$, is often used to characterize a spread spectrum transmission system.

The *processing gain*, sometimes called the *spreading factor,* is defined as the ratio of the information bit duration over the chip duration:

$$\text{Processing Gain (PG)} = \text{SF} = T_b / T_c \tag{7.1}$$

It represents the number of chips contained in one data bit. Higher Processing Gain (PG) means more spreading. High PG also means that more codes can be allocated on the same frequency channel.

In general, PG is defined as the SNR (signal to noise ratio) at the output to the SNR at the input of the signal processor. If the value is expressed in dB then

$$\text{PG (dB)} = \text{SNR}_{\text{out}}\text{ (dB)} - \text{SNR}_{\text{in}}\text{ (dB)} \tag{7.2}$$

Consider an example with a transmitting signal data rate of 9.6 kbps and a code rate of 4.2288 Mcps. The processing gain is 21 dB. If SNR_{out} is 7 dB then SNR_{in} is –14 dB below the interference signals caused by the sum of all signals, which appears as noise like to all other users except the required one.

7.5 SPECTRAL EFFICIENCY OF DIFFERENT WIRELESS ACCESS TECHNOLOGIES

The scarce radio resources for wireless communication systems need to be maximally utilized. There are several ways to improve the spectral utilization such as, frequency reuse, cell planning, channel assignment and multiple access technologies. But the spectrum of the wireless channel is always interference limited. The spectral efficiency of any system also depends on many factors like channel spacing in kHz, cell area in km^2, frequency reuse factor, modulation techniques and multiple access technology (FDMA, TDMA, CDMA, etc.).

There are two ways of representing spectral efficiency η—one, in terms of available data channel per unit bandwidth per unit coverage area (channels/MHz/km^2) and two, in terms of total data traffic per unit bandwidth per unit area (Erlangs/MHz/km^2). The second definition is more practical, because there is always data traffic blocking during the busy hours of a day.

$$\eta = \frac{\text{Total data traffic carried by the communication systems}}{\text{System bandwidth} \times \text{Coverage area}}$$

in (channels/MHz/km^2) (Definition 1)

$$\eta = \frac{\text{Total available data channels in the communication systems}}{\text{System bandwidth} \quad \text{Coverage area}}$$

in (Erlangs/MHz/km^2) (Definition 2)

In this section, the spectral efficiency of a wireless system in different access techniques are analyzed.

7.5.1 Spectral Efficiency in FDMA System

AMPS, the first analog system in cellular technology used in US, was based on the FDMA/FDD system. The single channel (which is basically two simplex channels) is assigned to each user during call progression. FDD is used to separate the forward and reverse channels by 45 MHz frequency as shown in Fig. 7.2. The channel is released for the use of another user after handover to the new base station. The total bandwidth BW_t is divided into a number of channels each with B_{ch} bandwidth and two guard bands at the ends as shown in Fig. 7.9. If B_G is the bandwidth of a guard band, then the number of available channels for an AMPS system is

$$N_{ch} = (BW_t - 2B_G) / B_{ch} \tag{7.3}$$

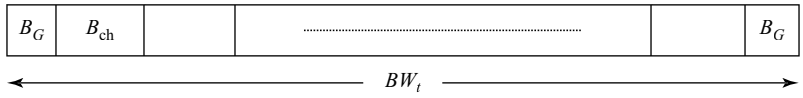


Fig. 7.9 *AMPS uplink or downlink channel structure*

In the definition of spectral efficiency, available data channel is the required parameter. The total N_{ch} is divided for data and control channels. So,

$$BW_t = N_{data}\, B_{ch} + N_{control}\, B_{ch} + 2B_G \tag{7.4}$$

The total geographical area is divided into cells. Cell clustering is made for frequency reuse to enhance system capacity. Considering one cell cluster into the system and the total BW_t is allocated to one cluster, the spectral efficiency for an FDMA system can be defined as

$$\begin{aligned} \eta_{FDMA} &= \text{Data channel available / Total system bandwidth} \\ &= N_{data}\, B_{ch} / BW_t < 1 \end{aligned} \tag{7.5}$$

If N is the number of cells per cluster and one cluster within the system is considered then Eq. (7.3) represents the number of channels/cluster ($N_{ch/cluster}$), from which some of the channels are available for data and some for control channels. So, the number of data channels cell ($N_{data/cell}$) can be expressed as

$$\begin{aligned} N_{data/cell} &= N_{data/cluster} / N = (N_{ch/cluster} - N_{control}) / N \\ &= \frac{\{(BW_t - 2B_G) / B_{ch}\} - N_{control}}{N} \end{aligned} \tag{7.6}$$

So, from the definition 1 of spectral efficiency, η can be expressed as

$$\begin{aligned} \eta &= \text{Number of data channels per cluster} / (\text{System bandwidth} \times \text{Area of the cluster}) \\ &= N_{data/cluster} / (BW_t \times N \times A_{cell}) \\ &= N \times N_{data/cell} / (BW_t \times N \times A_{cell}) \text{ channels/MHz/km}^2 \end{aligned} \tag{7.7}$$

where A_{cell} is the area of each cell. $N_{data/cell}$ is defined by Eq. (7.6). If the $N_{control/cell}$ is known with other defined system parameters for an FDMA system, then the spectral efficiency can be calculated from Eq. (7.7).

From the definition 2,

$$\eta = \text{Number of data traffic per cluster / (System bandwidth} \times \text{Area of the cluster)}$$
$$= \eta_{\text{trunk}} \times N_{\text{data/cluster}} / (BW_t \times N \times A_{\text{cell}})$$
$$= N \times \eta_{\text{trunk}} \times N_{\text{data/cell}} / (BW_t \times N \times A_{\text{cell}}) \text{ Earlangs/MHz/km}^2 \quad (7.8)$$

where η_{trunk} is the trunk efficiency in each cell that is related with the blocking probability of the data traffic and the value is less than one. Due to congestion at busy hours, all data traffic may not successfully get through the system. The actual value of spectral efficiency in terms of data traffic in Erlangs is less than the efficiency in terms of channels.

7.5.2 Spectral Efficiency in TDMA System

The efficiency of the TDMA system depends on percentage time slot used for data transmission within a frame. As stated earlier, TDMA may be wideband or narrowband as shown in Fig. 7.3. Calculation of spectral efficiency for the two types is given below.

Spectral Efficiency of WTDMA The entire bandwidth is used for each user at a defined time slot. So, the efficiency is the measure of the percentage time slot within a TDMA frame. With respect to Fig. 7.4, it is seen that time in one TDMA frame is divided for the preamble, trailer bits and information message. During data transmission, each user gets a particular time slot. Again, one slot time is divided for trail bits, sync bits, guard bits and the remaining part is used for information bits.

Let the time duration for preamble, trailer and one frame be defined as T_p, T_t and T_f respectively. Let in each time slot there be N_s number of symbols of which N_I is the number of information bits. Then the wideband TDMA efficiency can be defined as

$$\eta_{WTDMA} = \frac{(T_f - T_p - T_t)}{T_f} \times \frac{N_I}{N_s} \quad (7.9)$$

Spectral Efficiency of NTDMA For a narrowband TDMA system, the total bandwidth (BWt) of the system is divided into N sub-bands each having a bandwidth of B_{ch}, and there are guard bands (B_G) at the two ends similar to the AMPS FDMA system (as drawn in Fig. 7.9). Each channel (sub-band) is divided for slot times, preamble and trailer. The spectral efficiency of NTDMA is η_{WTDMA} multiplied by a factor, which is the ratio of the available information bandwidth ($N_{\text{sub}} \times B_{\text{ch}} = BW_t - 2B_G$) to the total system bandwidth (BWt) where Nsub is the number of sub-bands. So, the spectral efficiency of the NTDMA system is

$$\eta_{\text{NTDMA}} = \frac{(T_f - T_p - T_t)}{T_f} \times \frac{N_I}{N_s} \times \frac{N_{\text{sub}}\ B_{\text{ch}}}{BW_t} \quad (7.10)$$

The maximum number of users that can be supported in the NTDMA system is the number of sub-band times the number of slots = $N_{\text{sub}} \times N_{\text{slot}}$. For WTDMA, $N_{\text{sub}} = 1$, and N_{slot} corresponds to the number of users that can be supported within a channel.

$$N_{\text{user}} = \frac{N_{\text{slot}} \times (BW_t - 2B_G)}{B_{\text{ch}}} \quad \text{user/channel} \quad (7.11)$$

This is without considering the frequency-reuse factor. If N is the reuse factor for the TDMA system then the maximum number of users within a cell can be obtained by dividing Eq. (7.11) by N, the frequency reuse factor.

$$N_{\text{cell}} = N_{\text{user}} / N \quad \text{user/cell} \quad (7.12)$$

7.5.3 Spectral Efficiency for DS-CDMA System

The CDMA system is different from FDMA and TDMA. The entire frequency bandwidth is available for each user. The system capacity, or spectral utilisation factor, is defined by the maximum number of users that can be accommodated within the system maintaining a certain level of quality of service (QoS). The capacity in a CDMA system is interference limited whereas in TDMA and FDMA systems, it is bandwidth limited. The link performance is directly related with the number of users within the CDMA system. So, to see the spectral efficiency of a CDMA system, a different approach is taken. Some of the related parameters those affect the system performance is discussed first. Extraction of a signal at the receiver end in the presence of system noise and interference from the transmissions of all users and maintaining a certain level of signal-to-noise ratio in the desired signal is the objective of the receiver. To quantify the performance of the CDMA receiver, let us first describe some useful terms. The signal levels themselves are referred as a ratio to the level of the total noise power, which is denoted as E_b/N_o (bit energy per total noise power spectral density). Evaluation of Bit Error Rate (BER) with respect to E_b/N_o is the performance measure of the system. In a CDMA system, it is also needed to examine the baseband performance in terms of its chip energy and thus E_c/N_o is another quantity to measure.

Figure 7.10 is the illustration for the relation between chip rate and bit rate. The energy per information bit is defined as the energy carried by the RF carrier over an information bit period T_b. The variable E_b is used to represent this energy (= $A^2T_b/2$, where A is the signal amplitude).

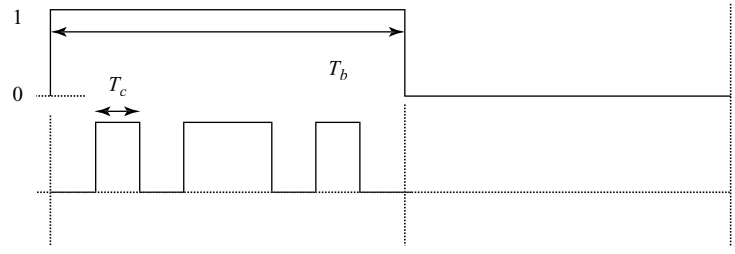


Fig. 7.10 *Relationship between chips and bits*

The noise power is referred as N_o, and it represents the noise power spectral density in terms of watts per Hz.

E_c is the energy per chip. In Fig. 7.10, there are 8 chips within a single information bit. So the spreading factor (SF) is taken as 8. The energy per chip is related to the energy per information bit via the SF ($E_b = E_c \times \text{SF}$). As described previously the processing gain (PG) is another important parameter in a CDMA system. It is the ratio between E_b/N_o to E_c/N_o . PG is same as the SF. If the PG of a system is greater than 1 then E_c/N_o for the system is negative. This implies that the desired signal is below the noise floor of the received signal.

In calculating the capacity of a CDMA system, first consider a cellular system consisting of a single base station with multiple mobile users simultaneously transmitting to the base station. The linear combiner at the base station adds all the spread signals received from the mobiles. In general, a base station controls the power for each of the received signals. To do this, it uses a weighting factor for each signal for forward-link power control. A mobile set uses the pilot signal to control the power level for the reverse link. In a single cell system with power control, it is considered that the base station receives the same power from all the mobiles in the reverse link.

As shown in Fig. 7.11, let there be N_m number of mobiles each transmitting to the base station at the cell site. Each demodulator at the base station receives the composite signal of all along with the desired signal. Let P_d be

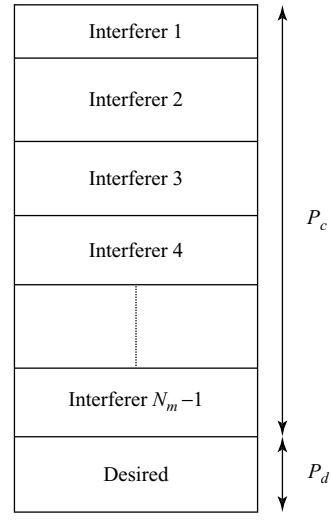


Fig. 7.11 *Uplink interferer signals*

the power of the desired signal and P_c be the power of the composite signal from all interferer signals of numbers $(N_m - 1)$. The ideal transmission power of BTS (base transmitting system) and the mobile terminal is achieved when the receiver is capable of de-modulating and de-spreading the information received within the frame error rate (FER) limit specified by the system while providing the lowest interference level to the other users.

$$\begin{aligned}&\text{Considering the equal power level of each signal} = S,\ P_c = (N_m - 1)S \text{ and } P_d = S\\ &\text{Then signal-to-interference ratio} = S/N = P_d/P_c = 1/(N_m - 1)\end{aligned} \tag{7.13}$$

The CDMA system is designed in a way that all channels transmitted by several users within a cell reach the BTS with a sufficient E_b/N_o that provides acceptable FER (frame error rate).

If S is the power assigned to the communication channel in watts, T_b is the duration of information bits in seconds, R_b is the information transmission rate in kbps, N is the total noise power and W is the spectrum bandwidth in MHz then

$$\begin{aligned}E_b/N_0 &= (T_b \times S)/(N/W) = (1/R_b) \times (S/N) \times (W)\\ &= (S/N) \times (W/R_b) = (S/N) \times \text{PG}\\ &= (W/R_b)/(N_m - 1)\end{aligned} \tag{7.14}$$

where PG is the processing gain. The factors (S/N) and (E_b/N_o) are the quality of service factors at the BTS receiver.

Interference is a combination of intracell interference and intercell interference. In addition to that there is background noise. Within a single cell, the interference due to all other mobiles to a desired one is called *intercell interference*, whereas for a multicell system, there are other neighbor cells due to frequency reuse. Considering a hexagonal cellular array, there are $6i$ numbers of cells in each tier, $i = 1, 2, \ldots n$. For a CDMA system, frequency reuse is 1.

So, there will be a factor k that will be responsible for contributing both intra and intercell interference and needs to be multiplied to get total interference power as $[(N_m - 1) \times S] \times k$. Again, considering the system background noise with power P_n, the S/N ratio can be modified as

$$S/N = 1/\left((N_m - 1) \times k + P_n/S\right) \tag{7.15}$$

Defining $1/k = \eta_f$ the factor related with frequency reuse efficiency in a multicellular CDMA system which is less than one ($\eta_f = 1$ for single cell), the E_b/N_o ratio can be represented as

$$\begin{aligned}E_b/N_o &= (S/N) \times (W/R_b) = (S/N) \times \text{PG}\\ &= \text{PG}/((N_m - 1) \times k + P_n/S)\\ &= \frac{PG \times \eta_f}{(N_m - 1) + \eta_f (P_n/S)}\end{aligned} \tag{7.16}$$

So, the number of users that can be supported in a DS-CDMA system is

$$N_m = 1 + \eta_f[\ \text{PG}/(E_b/N_o) - P_n/S] \tag{7.17}$$

In deriving Eq. (7.17), we have considered perfect power control by the base station, omni-directional antenna structure at the BTS and equal call handling of all cells.

Antenna sectorization can be done in reducing interference. For three sectors 1/3 of interference can be reduced than with the omnidirectional antenna (as for three sectors, interference is 1/3 of the interference than that the omnidirectional antenna and it is proved in Chapter 3). Again by controlling voice activity, i.e., switching off the transmitter during the no-transmission phase of voice, the interference level can be reduced as the number of interferers is reduced. Let the new number of users in each sector be N_s in a single-cell system, and β_v is the parameter to account for voice activity factor. Then the ratio E_b/N_o will

be modified with the new quantity E_b/N'_o, where N'_o is the reduced noise level. The maximum number of mobile users within a sector can be redefined as

$$N_{\text{sector}} = 1 + (1/\beta_v)\,[\text{PG}/(E_b/N'_o)] \tag{7.18}$$

To get this relation, we consider that interference noise is predominant than the system background noise. So the term (P_n/S) is neglected. Hence, using antenna sectorization and controlling voice activity factors, the maximum number of users within a cell in a CDMA system can be increased. This implies the increase of the spectral utilization factor. To get this number, the QoS parameter is the E_b/N_o ratio that has to be maintained below a threshold limit.

So, the system utilization U_s is defined as the number of users that can be supported under the constraint that $E_b/N_o >= (E_b/N'_o)$. The overall capacity depends on many factors like power control error, soft handoff, antenna sectorization, etc.

The spectral efficiency of the DS-CDMA system is defined as a unitless quantity or in terms of bits/s/Hz as

$$\eta_{\text{DS-CDMA}} = (E_b/N'_o)/(E_b/N_o) \text{ where } E_b/N_o >= (E_b/N'_o) \text{ or}$$
$$\eta_{\text{DS-CDMA}} = U_s \times R_b/W \text{ bits/s/Hz}$$

where R_b and W are the information bit rate and one-way system bandwidth in Hz, respectively.

Example 7.1 The total bandwidth in an AMPS cellular system is allocated as 12.5 MHz. Using FDMA, 416 numbers of available channels with a spacing of 30 kHz are allocated to the users.

(a) What is the guard bandwidth used in the system?
(b) What is the spectral efficiency for this system if there are 21 channels used for control signaling?
(c) If the cell area is 8 km^2 and the frequency reuse factor is 4, find the overall spectral efficiency of the system.
(d) If the trunk efficiency of the system is 0.9, what is the spectral efficiency in Earlangs/MHz/km^2.

Solution

(a) Let B_G be the bandwidth for the guard band. For an FDMA system, the total bandwidth B_t is allocated among N number of users with channel bandwidth B_{ch}. There are two guard bands at the ends So,

$$N \times B_{\text{ch}} = B_t - 2B_G$$
$$B_G = (B_t - N \times B_{\text{ch}})/2 = (12.5 \times 10^6 - 416 \times 30 \times 10^3)/2 = 10 \times 10^3$$
So, $B_G = 10$ kHz

(b) η_{FDMA} = Data channel available / Total system bandwidth
$= N_{\text{data}}\, B_{\text{ch}}/B_t = 30 \times 10^3 \times (416 - 21)/(12.5 \times 10^6)$
$= 30 \times 395/(12.5 \times 10^3) = 0.948$

So, the efficiency of the FDMA-based AMPS without frequency reuse is 94.8%.

(c) The number of available data channels per cell = (416 – 21)/4 = 395/4 = 98.75

$$\text{The overall spectral efficiency } \eta_{\text{channel}} = \frac{\text{Available data channel per cell}}{\text{Total bandwidth} \times \text{Area of the cell}}$$

$$= 98.75/(12.5 \times 8) = 0.9875 \text{ channels/MHz/km}^2$$

(d) η_{Earlang} = Trunk efficiency $\times\ \eta_{\text{channels}}$
$= 0.9875 \times 0.9 = 0.88875$ Earlangs/MHz/km^2

Example 7.2 In a GSM system with a 25-MHz forward link, there are 200 kHz radio channels allocated for voice communication using TDMA/FDD, and each channel can support 8 simultaneous speech channels each with a time slot of 0.577 ms.

(a) What are the total numbers of users that can be supported?
(b) What is the duration of a frame?
(c) What is the time gap between two successive transmissions for a particular user?

Solution

(a) A 25-MHz forward link is allocated to 200-kHz channels. Each channel can support 8 simultaneous users using TDMA. So, the total number of users N_{GSM} that can be accommodated in the GSM system is
$N_{GSM} = 25 \times 10^6 / (200 \times 10^3/8) = 1000$

(b) The time duration of a slot $T_s = 0.577$ ms
So, time duration of a frame $T_f = 8 \times T_s = 0.577 \times 8 = 4.616$ ms

(c) We know that in a GSM system, each user gets one time slot in each frame using the TDMA technique. So, a particular user has to wait for a frame-duration time for getting a slot for the next transmission. The time between two successive transmissions is 4.616 ms.

Example 7.3 In the GSM cellular system defined in Problem 7.2, if the frequency reuse factor is 7 then (a) what will be the total number of users per cell? Now suppose β_v is the parameter to account for the voice-activity factor, defined as the percentage of time that an active mobile user is transmitting data. There may be alternate on and off state during voice communication, and during the off period other users can be accommodated for transmission in the asynchronous mode of TDMA. (b) If β_v is 50%, what is the effective cell capacity?

Solution

(a) In Problem 7.2, let us assume that the entire capacity is assigned to a cell cluster. So,
$N_{GSM} = 1000$ is the number of users/cluster without frequency reuse.
Now $N = 7$, so $N_{user/cell} = 1000/N = 1000/7 = 142.8 \cong 143$ users/cell.

(b) Considering the voice activity factor $\beta_v = 50\% = 0.5$, the effective cell capacity can be increased as the off transmission time is used for other active users. So, with β_v, the number of users $N_{v\text{-act}}$ per cell is
$N_{v\text{-act}} = N_{user/cell} / \beta_v = 1000/(N \times 0.5) = 286$ users/cell.

Example 7.4 Consider a GSM cellular system with a frequency reuse factor of 7. It has the uplink and forward links, each having 25-MHz bandwidths. There are 125 duplex channels, each having a bandwidth of 200 kHz with 45 MHz frequency separation. Each channel supports 8 users using TDMA with a slot duration of 0.577 ms and a frame duration of 4.615 ms. Using GMSK modulation, the bandwidth efficiency is achieved as 1.25 bits/s/Hz. The speech transmission rate is 13 kbps, and the channel-coding results in a coded bit rate as 22.8 kbps. Consider only a normal TDMA frame for speech transmission with 8 time slots. If each slot has 156.25 bits, 3 start bits, 116 coded speech bits, 26 training bits, 3 stop bits and 8.25 guard bits, determine

(a) the spectral efficiency for narrowband TDMA for this system, and
(b) the overall spectral efficiency in bits/s/Hz/cell.
(c) If the system uses a guard band of 20 kHz then repeat the problem for parts (a) and (b).
(d) Repeat part (c) to calculate spectral efficiency for a frequency reuse of 4 and 9.

Solution:

(a) From Eq. (7.10), we can write the efficiency for NTDMA as

$$\eta_{\text{NTDMA}} = \frac{(T_f - T_p - T_t)}{T_f} \times \frac{N_l}{N_s} \times \frac{N_{\text{sub}} B_{\text{ch}}}{BW_t}$$

$$= \frac{(T_f - T_p - T_t)}{Tf} \times \frac{N_1}{N_s} \times \frac{BW_t - 2B_G}{BW_t}$$

As there are 125 duplex channels = 25 MHz / 200 kHz = 125, guard band $B_G = 0$

$\eta_{\text{NTDMA}} = (4.615 - 0 - 0)/4.615 \times (116/156.25) \times (25 - 2 \times 0)/25$

$= 0.7424$

(b) Given the bandwidth efficiency = 1.25 bps/Hz, overall efficiency factor due to modulation and channel coding = $\varepsilon = 1.25 \times$ (transmission rate/coded bit rate)

$\eta_7 = (\varepsilon \times \eta_{\text{NTDMA}})$/Frequency reuse factor

$= 1.25 \times (13/22.8) \times 0.7424 \times 1/7 = 0.075588$ bits/s/Hz/cell

(c) $\eta_{\text{NTDMA}} = (4.615 - 0 - 0)/4.615 \times (116/156.25) \times (25000 - 20)/25000$

$= 0.7364$

$\eta = (\varepsilon \times \eta_{\text{NTDMA}})$/Frequency reuse factor

$= 1.25 \times (13/22.8) \times 0.7364 \times 1/7 = 0.074984$ bits/s/Hz/cell

(d) $\eta_4 = (\varepsilon \times \eta_{\text{NTDMA}})$/Frequency reuse factor

$= 1.25 \times (13/22.8) \times 0.7364 \times 1/4 = 0.1312$ bits/s/Hz/cell

$\eta_9 = (\varepsilon \times \eta_{\text{NTDMA}})$/Frequency reuse factor

$= 1.25 \times (13/22.8) \times 0.7364 \times 1/9 = 0.0583$ bits/s/Hz/cell

Example 7.5 In a GSM system with the parameters given in Problem 7.4, the total service area is $A_s = 4500$ km^2 and the area of cell $A_{\text{cell}} = 9$ km^2. During busy hours, on an average, each user makes 2 calls having a duration of 5 minutes. The trunking efficiency of the system is 0.85. Determine

(a) the total number of cells in the system,
(b) the number of calls per hour per cell,
(c) the average number of users served per hour per cell, and
(d) the overall system efficiency in Earlangs/MHz/km^2.

Solution

(a) Number of cells = $A_s/A_{\text{cell}} = 4500/9 = 500$

(b) The number of data channel per cluster = $25000 \times 8/(200) = 1000$
Number of data channel per cell = $1000/7 = 142.8 = 142$
Number of calls per hour per cell = $142 \times (60/5) \times 0.85 \approx 1457$ calls/hour/cell

(c) The average number of users served per hour = Number of calls per hour per cell / average number of calls per user per hour = $1457/2 \approx 728$ users/hour/cell

(d) The number of available data channels per cell = 142
Spectral efficiency η = Data channel per cell/ (Cell area × Total bandwidth)
$= 142/(9 \times 25)$
$= 0.6311$ channels/MHz/km^2
The overall spectral efficiency of the system in Earlangs/MHz/km^2
= Trunking efficiency × Efficiency in channels/MHz/km^2
$= 0.85 \times 0.6311$
$= 0.53644$ Erlangs/MHz/km^2

Example 7.6 In a CDMA cellular system with an omnidirectional antenna, the required E_b/N_o ratio is 10 dB. If 250 users, each with a base-band data rate of 13 kbps are to be accommodated, determine the minimum channel chip rate of the spread spectrum

sequence for the two cases:

(a) Without considering voice activity factor.
(b) Considering a voice activity factor equal to 0.4.

Solution

(a) *Eb/No =10 dB = 10* *Nm = 250, Tb = 1/13 × 103 Tc = ?*
Without voice activity factor $E_b/N_o = \text{PG}/(N_m - 1)$
$10 \times (N_\text{m} - 1) = (1/13) \times 10^{-3}/T_c$
$T_c = (1/13) \times 10^{-3}/2490$
$= 3.237 \times 10^7$ chips/s

(b) With voice activity factor
$N_m = 1 + (1/\beta_v)\,[\text{PG}/(E_b/N_o)]$ and $\beta_v = 0.4$
$T_c = 0.4 \times 3.237 \times 10^7$ chips/s $= 1.2948 \times 10^7$ chips/s

Example 7.7 A CDMA system is defined with the following parameters:

Frequency reuse efficiency $\eta_f = 0.55$, $E_b/N_o = 10$ dB, the information bit transmission rate is 16.2 kbps, system bandwidth W = 12.5 MHz. Neglecting all other sources of interference, determine the system capacity and spectral efficiency of the CDMA system.

Solution

We know for a CDMA system,

$$E_b / N_o = \frac{PG \times \eta_f}{N_m - 1 + \eta_f (P_n / S)}$$

PG is the processing gain = W/R_b . Neglecting P_n/S

$$E_b / N_o = \frac{P_G \times \eta_f}{N_m - 1}$$

System capacity $N_m = [\text{PG} \times \eta_f / (E_b/N_o)] + 1 = [(W/R_b) \times \eta_f / (E_b/N_o)] + 1$
$= [(12.5 \times 10^6/16.2 \times 10^3) \times 0.55/10] + 1$
$= 43$ users/cell
$\eta_{\text{DS-CDMA}} = N_m \times (R_b/W)$ bits/s/Hz
$= 43 \times 16.2/(12.5 \times 10^3) = 0.0557$ bits/s/Hz

Example 7.8 A cellular system has a total bandwidth of 25 MHz. Each full duplex voice or control channel uses two 25 kHz simplex channels. It is assumed that the system uses a 7-cell reuse pattern, and 0.40 MHz of the total bandwidth is allocated for control channels. The system service area consists of 30 cells each having an area of 5 km^2 and a call blocking probability of 2%, as given by the Erlang - B formula. If the offered traffic per user is 0.025 Erlang then calculate

(a) the number of voice channels per cell,
(b) the trunking efficiency,
(c) the total number of users in each cell and in the system,
(d) the number of mobile users per channel in each cell and in the system,
(e) the maximum number of users in service at any instant in the system, and
(f) the spectral efficiency of the system in channels/MHz/km^2 and in Earlangs/MHz/km^2.

Solution

(a) Given, total bandwidth = 25 MHz = 25×10^3 kHz.
Individual channel bandwidth = 25×2 kHz = 50 kHz/duplex channel.
Total number of channels = $(25 \times 10^3)/ 50 = 500$ channels
Number of control channels = (400/50) = 20 channels
Number of available voice channels = (500 – 20) = 480 channels
Number of total cells in the system = 30 cells and the cluster size $N = 7$.
$\therefore$ number of voice channels per cell = (480/7) = 68.57

(b) Blocking Probability = 2%.
$\therefore$ from Erlang–B formula, we have for 68 channels and 2% of blocking probability, the traffic load per cell = 57.226 Erlangs
$\therefore$ trunking efficiency (η_t) = (57.226/68) = 84.15%

(c) The total number of users in each cell = (Total load traffic)/ (Offered traffic per user)
= (57.226 E/0.025 E) = 2289
If the system has 30 cells, the total number of users
= $2289 \times 30 = 68670$

(d) The load per channel = (57.226/68) E = 0.8415 E
Offered traffic per user = 0.025 E.
Number of users per channel in each cell = (0.8415/0.025) = 33.66
In the system, the number of mobile user per channel = $(33.66 \times 30) = 1009.8 = 1009$

(e) The maximum number of users in service at any instant in the system = $(68 \times 30) = 2040$

(f) The spectral efficiency
η_{Channels} = Number of data channels per cell / (Total bandwidh × Area of the cell) channels/MHz/km^2 = $68/(25 \times 5) = 0.544$ channels/MHz/km^2
η_{Earlang} = Trunking efficiency × number of data channels per cell /(Total bandwidth × Area of the cell) Erlangs/MHz/km^2 = $0.8415 \times 68 / (25 \times 5) = 0.45776$ Erlangs/MHz/km^2

Summary

Wireless communication systems are characterized by different multiple access techniques in the uplink to share the available radio spectrum among number of users and increase the spectral efficiency. The most commonly used multiple access techniques are FDMA, TDMA, CDMA and SDMA, where each system has its own pros and cons. FDMA is a narrowband mainly employed in first-generation cellular systems with low data rate. The second-generation cellular system is digital and TDMA is used for these systems. CDMA is an interference-limited multiple access technique mainly targeted for third generation cellular systems to support multimedia services, though it is also used for second-generation IS-95 systems. The details of these techniques with their advantages and disadvantages have been discussed. Also, the spectral efficiency of each technique has been derived. The numerical solutions of different problems are really helpful for complete understanding of these access technologies.

References

[1] Viterbi, A.J., *CDMA-Principle of Spread Spectrum Communication,* Addison–Wesley, 1995.
[2] Sheen, W., and G.L. Stuber, *Effects of Multipath Fading on Delay-Locked Loops for Spread Spectrum Systems,* IEEE Trans. on Communications, 42(2/3/4): pp.1947–1956, Feb/Mar/Apr 1994.
[3] Sourour, E.A., and S.C. Gupta, *Direct Sequence Spread Spectrum Parallel Acquisition in a Fading Mobile Channel,*IEEE Trans. on communication, 38(7): pp. 992–998, July 1990.
[4] Korowajczuk, L., *et.al*, *Designing CDMA2000 Systems*, John Wiley, West Sussex, England, 2004.
[5] Mark, J.W., and W. Zhuang, *Wireless Communication and Networking,* PHI, 2003.
[6] Kim, K.I., *Handbook of CDMA System Design, Engineering and Optimization,* Prentice Hall, Upper Saddel River, NJ, 2000.
[7] Lee, W.C.Y., "Overview of Cellular CDMA", *IEEE Trans. on Vehicular Technology*, Vol. 40, No. 2, May 1994.
[8] Gudmundson, B., J. Skold, and J.K. Ugland, A Comparison of CDMA and TDMA Systems, Proc. of the *42nd IEEE Vehicular Technology Conference*, Vol. 2, pp. 732–735, 1992.
[9] Rappaport, T.S., and L.B. Milstein, "Effects of Radio Propagation Path Loss on DS-CDMA Cellular Frequency Reuse Efficiency for Reverse Channel", *IEEE Trans. on Vehicular Technology*, Vol. 41, No.3, pp. 231–242, Aug 1992.
[10] Raith, K., and J. Uddenfeldt, "Capacity of Digital Cellular CDMA Systems", *IEEE Trans. on Vehicular Technology*, Vol 40, pp. 323–331, May 1991.
[11] Pickholtz, R.L., L.B. Milstein, and D. Schiling, "Spread Spectrum for Mobile Communications", *IEEE Trans. on Vehicular Technology*, Vol 40, pp. 313–322, May 1991.

Questions for Self-Test

7.1 Duplexing is required for two-way simultaneous communication.
(a) False (b) True
7.2 FDD provides two separate bands of frequencies for any user.
(a) False (b) True
7.3 In an FDMA system, a duplexer is used only in a user handset, but not in a base station.
(a) True (b) False
7.4 Frequency separation is a must between uplink and downlink.
(a) False (b) True
7.5 FDMA is suitable for multimedia communication.
(a) True (b) False
7.6 Inter-modulation occurs due to non-linearity effect in FDMA system.
(a) True (b) False
7.7 TDMA needs synchronization among users.
(a) False (b) True
7.8 TDMA uses only TDD for the uplink and downlink separation.
(a) True (b) False
7.9 A single radio channel in a TDMA system means
(a) an assigned time slot (b) a frequency channel
7.10 In asynchronous TDMA, the frame length is
(a) variable (b) fixed
7.11 Data transmission in TDMA system occurs
(a) in bursts (b) continuously

7.12 TDMA systems use _____ for extracting timing information
(a) matched filter (b) co-relation detector (c) both a and b

7.13 TDMA helps for easy handoff because of _____ data transmission.

7.14 In CDMA, signal separation is achieved by the _____ property of the codes.

7.15 In spread spectrum, the base-band signal is modulated with a signal having _____ bandwidth.

7.16 For the following systems, identify which are analog, digital, FDMA, TDMA and CDMA based. AMPS, GSM, PDC, IS-95, DECT

7.17 What is the frequency of separation between uplink and downlink in AMPS and GSM?

7.18 Indicate true or false.
(a) There are 8 half-rate users in one time slot in GSM.
(b) GSM uses slow frequency hopping.
(c) GSM supports soft handoff.

7.19 Which cellular system among the following solely supports FDMA and TDMA?
DECT, CT2, USDC, GSM

7.20 How many number of full-rate voice channels per cell are available in a GSM system?

7.21 Why are guard times used in TDMA frame?

7.22 What is the main limitation of SDMA?

7.23 On what factors does the spectral efficiency of a wireless system depend? Give two definitions for spectral efficiency.

7.24 What is called processing gain in CDMA systems? What is its significance?

7.25 What would happen if there were no power control mechanisms in CDMA systems?

7.26 What is the relationship between the maximum number of users and E_b/N_o ratio in any CDMA system? Discuss the role of E_b/N_o in computing spectral efficiency.

7.27 How does antenna sectorization and voice activity factor influence the maximum number of users in a CDMA system?

7.28 Derive the spectral efficiency for a narrowband TDMA system. What is the maximum numbers of users that can be supported within a cell?

7.29 Discuss how spectrum allocation is done in an AMPS cellular system?

7.30 Discuss how spectrum allocation is done in a GSM cellular system?

7.31 How does capacity and spectral efficiency of a cellular system depend on the frequency reuse factor?

7.32 How does area of a cell affect the spectral efficiency of a cellular system?

7.33 What is trunking efficiency? Does it play any role in the spectral efficiency of a wireless system?

7.34 How many full-rate voice channels per cell are available in an AMPS system with a total bandwidth of 25 MHz if the guard band is 10 kHz and the channel spacing is 30 kHz?

Ans. 832

7.35 Consider an FDMA based AMPS network with a total bandwidth of 12.5 MHz, a channel bandwidth of 30 kHz, and a guard bandwidth of 10 kHz. How many users can be accommodated? Derive the expression for spectral efficiency of the FDMA system. What is the upper bound of the efficiency?

Ans. 0.9984

7.36 In Problem 7.35, if there are 21 control channels, the frequency reuse factor is 7, the area of each cell is 6 km^2, the total service area is 3000 km^2, the average number of calls per user during busy hours is 4.2, the average call holding time is 100 s, and the call blocking rate is 2% then calculate
(a) number of calls/hour/cell
(b) number of calls/hour/km^2
(c) number of users/hour/cell
(d) number of users/hour/channel
(e) the overall spectral efficiency of the system

Ans. a.1651 b. 275, c. 393 d. 29 e. 0.6113 Earlangs/MHz/km^2

7.37 What is the duration of a bit in a GSM system? If 8 voice channels are supported in each radio channel and there are no guard bands, then how many simultaneous users can be accommodated in a GSM system? ***Ans.*** 1000

7.38 If there are 3 trailing bits, 3 stealing bits, 26 training bits, 8.25 guard bits and the rest 116 are data bits, what is the percentage overhead in a GSM frame? If the GSM frame is 4.615 ms, what is the spectral efficiency for wideband TDMA? ***Ans.*** 0.7424

7.39 Consider a GSM cellular system with a frequency reuse factor of 9. It has the uplink and forward links, each having 25 MHz bandwidths. There are 125 duplex channels, each having a bandwidth of 200 kHz with 45 MHz frequency separation. Now, consider that the system bandwidth is divided into a number of sub-bands with a band gap of 20 kHz, and each sub-band is partitioned into time slots. Each channel supports 8 users using TDMA with a slot duration of 0.577 ms and a frame duration of 4.615 ms. Calculate

(a) the number of sub-bands,
(b) the maximum number of simultaneous users that can be accommodated during one use of available frequency spectrum,
(c) the number of users per cell, and
(d) the spectral efficiency for narrowband TDMA.

Ans. a. 124 b. 992 c. 110 d. 0.7412

7.40 In Problem 7.39, now consider that voice transmission occurs in alternate talk state (on state). If 60% of the time is the off state when other users use the spectrum, what will be the modified cell capacity of the GSM system?

Ans. 275 users/cell

7.41 Consider a GSM cellular system with a frequency reuse factor of 7. It has the uplink and forward links each having 25-MHz bandwidths. There are 124 duplex channels each having a bandwidth of 200 kHz with 45-MHz frequency separation. The speech transmission rate is 13.2 kbps and channel-coding results in a coded bit rate as 23.2 kbps. There are 8 users per radio channel, each using a time slot 0.577 ms within a frame. The ratio of control overhead to data is 0.742.

(a) What is the bandwidth of the guard band?
(b) What is the modulation efficiency?
(c) What is the effective number of users/cell?
(d) What is the spectral efficiency in bits/Hz/cell?

Ans. a. 200 kHz b. 0.5689 c. 142 d. 0.0598 bits/Hz/cell

7.42 An AMPS cellular system with a frequency reuse factor of 4 and a total bandwidth of 30 MHz uses two simplex 25 kHz channels for voice and control signals. Assume that a 1 MHz bandwidth is used for the control channel and only one control channel in each cell is needed. Calculate

(a) the number of users per cell,
(b) the number of voice channels in each cell, and
(c) calculate the above two parameters as in (a) and (b) if the frequency reuse factor changes to 7.

Ans. a 145 b.82

7.43 A CDMA system is defined with the following parameters:

7.44 Frequency reuse $\eta_f = 0.65$, Eb/ No = 12 dB, the information bit transmission rate is 19.2 kbps, system bandwidth W = 12.5 MHz. Neglecting all other sources of interference, determine the system capacity and spectral efficiency of the CDMA system.

Ans. 27 users/cell , efficiency = 0.041472 bits/s/Hz

7.45 In Problem 7.43, if Eb/ No now becomes 11 dB then what will be the spectral efficiency? Discuss the effect of Eb/ No on the system.

Ans. 34, and 0.05222 bits/s/Hz

7.46 In a single-cell CDMA system with an omnidirectional antenna, the Eb/ No = 10 dB is required for each user. If there are 150 users each with a transmission data rate of 13 kbps to be accommodated, what will be the minimum channel bit rate of the spread spectrum chip sequence considering no voice activity factor and with a voice activity factor of 0.4.

Ans. a. 1.9372 × 107 chips/s b. 0.7748 × 107 chips/s

7.47 Repeat Problem 7.45 with a 3-sector antenna system and a voice activity factor of 0.4. Discuss the result in comparison with Problem 7.45.

Ans. 2.3245 × 107 chips/s

Second-Generation Mobile Networks—GSM: Architecture and Protocols

8

Introduction

Global System for Mobile Communications (GSM) is a 2G digital mobile cellular system with its own communication protocols, interfaces and functional entities. Originally, GSM stood for 'Group Special Mobile' and intended to be a new telecommunication standard in Europe. The European Telecommunication Standards Institute (ETSI) was founded in 1988 and was responsible for standardizing the GSM technical specifications over Europe. The success of this standard has necessitated its renaming to Global System for Mobile Communications, which reflects its application worldwide. The first commercial GSM phone call was made in Finland on the 1st of July 1991.

The popularity and success of GSM can be summarized as follows:

- It is an open system standard, and anyone can have access to the specifications.
- It can give multi-vendor solutions because of a standardized interface.
- It can support roaming to users.
- It can maintain security and privacy of speech and data due to encrypted transmission.
- Digital transmission gives high speech quality and increases radio spectrum efficiency by the use of multiple access technology—FDMA, TDMA.
- Above all, it has an evolutionary implementation concept—upgradeable with downward compatibility.

ETSI has been standardized to operate on three basic frequency regions—900 MHz, 1800 MHz and 1900 MHz.

8.1 GSM NETWORK ARCHITECTURE

The GSM network architecture is given in Fig. 8.1. It consists of several base transceiver stations (BTS), which are clustered together and connected to a base station controller (BSC). Several BSCs are then

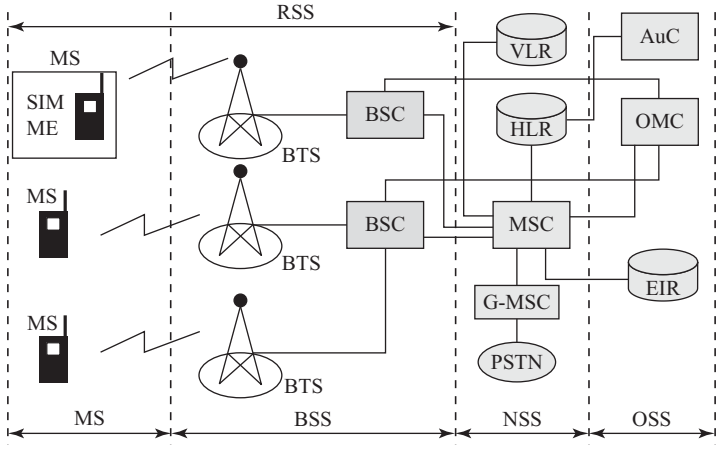


Fig. 8.1 *GSM network architecture*

connected to an MSC. The MSC has access to several databases, including the Visiting Location Register (VLR), Home Location Register (HLR), and Equipment Identity Register (EIR). MSC is responsible for establishing, managing, and clearing connections, as well as routing calls to the proper radio cell. It supports call rerouting at times of mobility. A gateway MSC provides an interface to the public telephone network.

All the identity information of the users such as home subscription base and service profiles are kept in HLR, which takes on a major role for mobile user tracking, while roaming from the home network to visited networks. The VLR stores information about subscribers visiting a particular area within the control of a specific MSC. The various functional blocks of a GSM network are explained in the following sections.

The three main sections of the GSM architecture are:

1. Radio Subsystem (RSS)
2. Network and Switching Subsystem (NSS)
3. Operation Subsystem (OSS)

8.1.1 Radio Subsystem (RSS)

Figure 8.2 shows the different entities for RSS. It consists mainly of Mobile Station (MS) and Base Station Subsystem (BSS). The MS consists of two major entities—Mobile Equipment (ME), the actual device and a smart card, and the Subscriber Identity Module (SIM), irrespective of whether the BSS has a Base Transceiver Station (BTS), a Base Station Controller (BSC) or a Trans coding and Rate Adaptation Unit (TRAU). In the following section, the functionalities of each entity will be discussed briefly.

SIM basically supports personal mobility; a user can use the SIM card in any mobile device without informing the network operators. The International Subscriber Identity Module (IMSI) is the unique identity number residing in SIM. A subscriber is identified with this number and along with the secret key used for authentication. A unique number called IMEI, International Mobile Equipment Identity, identifies the ME. These two numbers, IMSI and IMEI, are independent. This is because subscribers are not dependent on the mobile devices. Instead of using IMSI, another important number TMSI, Temporary IMSI, is often used as a user identifier while roaming and is periodically changed. This makes for tight user security by protecting the user from undesirable snooping. IMSI consists of the Mobile Country Code, Mobile Network Code, and Mobile Subscriber Identification Code.

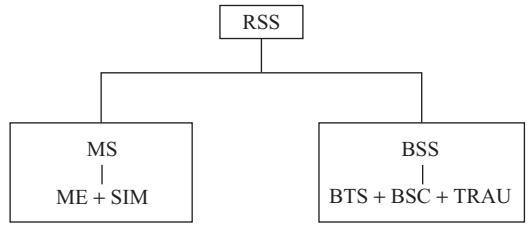


Fig. 8.2 *Radio subsystem of GSM*

BTS It is basically a radio transceiver system within the coverage area of a cell. It allows the MS to communicate with the network through a radio link. Transmission and reception at the BTS with the MS is done via an omnidirectional or directional antenna. The major functions of BTS are the transmission of the signals in a desired format. It does the coding and decoding, nullifies the propagation effects, like multipath, with the enhanced use of equalizer techniques. The data sent by the mobile is encrypted by BTS depending on the security information obtained from the network. In GSM, it can handle seven users at a time, with one channel being reserved for downlink broadcasting. By the use of frequency reuse and increased number of BTS, the system capacity can be increased. That means a greater number of users can be handled simultaneously.

BSC More than one BTS are connected to a BSC. A radio channel is set up by the BTS which handles the switchover of a mobile user from one BTS to another during roaming and also controls the handover process. The major functions of BSC are radio resource management and handover. It also controls the transmitted power. A BSC communicates to a BTS through Time Division Multiplex (TDM) over an A_{bis} interface.

Transcoding and Rate Adaptation Unit (TRAU) TRAU is the logical part of BSS, but resides close to MSC to reduce the significant transmission cost. The main communication in 2G systems is the voice, which is converted into a binary stream through a complex process. The voice data is sent in a 16 kbps channel through BTS and BSC to TRAU from the mobile unit. The most important role of TRAU is to convert this speech to 64 kbps rates over the PSTN or ISDN (Integrated Services Digital Networks). The voice traffic channel is 13 kbps and data traffic is 9.6 to 14.4 kbps on the air interface, hence through multiplexing and transcoding, it is converted to 64 kbps [9]. This is illustrated in Fig. 8.3.

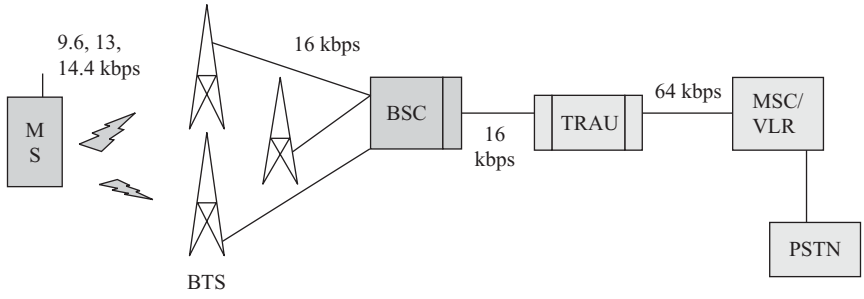


Fig. 8.3 *Example for transcoding*

8.1.2 Network Switching Subsystem (NSS)

The main basic components of NSS are the Mobile Switching Centre (MSC), Home Location Register (HLR) and Visitor Location Register (VLR) as shown in Fig. 8.4.

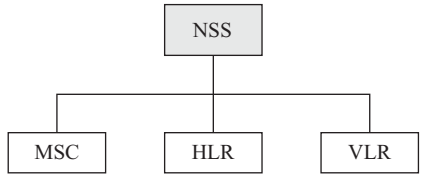


Fig. 8.4 *Components of NSS*

MSC In addition to the normal switching function like any other PSTN or ISDN, MSC plays a major role in supporting mobility of a user. It controls both switching and management together by controlling the number of BSCs attached to it. The MSC processes the request for the connection of a mobile user to the network and also deals with registration and authentication. When a request comes, MSC sends the request to the authentication center for the user information and performs authentication. Only then does the MSC register the MS with its associated VLR (sometimes co-located), in the visiting network. This location information is updated with the HLR that may reside in the same network or any other network while roaming the MS. The MSC also routes the call to and from the MS. There is a gateway MSC to connect the call to any other fixed network like PSTN or ISDN.

HLR The HLR maintains all the information related to a mobile subscriber in its database. The database in HLR remains intact and unchanged until the termination of the subscription. When a subscriber registers with any mobile network operator, a service level agreement (SLA) is formed. The operator's network is known as the home network. HLR is a huge database located within this home network which stores the administrative information of the mobile subscriber like IMSI, type of subscription, service that the user can get, subscriber current location, i.e., to which VLR, service restriction, and supplementary services of all home network (HN) subscribers. HLR also receives the connection information when the mobile subscriber registers to a different operator's network in a different country while traveling. For this, it requires roaming agreement between the two operators' networks.

VLR VLR is another temporary database to which the subscriber currently registers, and within its vicinity, it covers the service area of its associated MSC. VLR is dynamic in nature and interacts with the HLR when recording the data of a particular mobile subscriber. When a mobile user enters into a new service area, it

sends a request to the MSC for connection. After authentication, the MSC updates the VLR. This update information is then sent to the HLR. If a user under VLR makes a call then VLR precisely knows the position of the user in a single cell. Otherwise the position of the user will be known by the VLR in the location area (LA). LA is basically a group of cells connected to a single MSC within the service area.

8.1.3 Operation Subsystem (OSS)

OSS consists of the Authentication Center (AuC), Operating and Maintenance Center (OMC) and Equipment Identity Register (EIR) as given in Fig. 8.5.

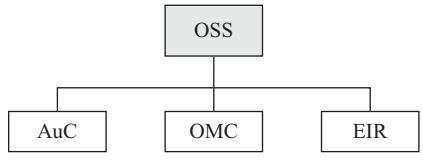


Fig. 8.5 *Components of OSS*

AuC AuC is a database that stores a copy of the secret key of the user's SIM card to enable authentication and encryption over the radio link. AuC has the required data to protect the network from false users and to protect the calls of the regular user. There are two secret keys in GSM standard: the encryption of communications between mobile users and the authentication of the users. These keys are kept both in the mobile equipment and the AuC. So the information is protected against any snooping access.

OMC OMC centrally monitors and controls the network elements for smooth running of the network and guarantees the best possible service quality for a network. For this purpose, it has to perform network monitoring, network development, network measurements and fault management.

EIR It stores International Mobile Equipment Identity (IMEI) of all valid mobiles on the network. This number is installed during the manufacture of the equipments and specifies the standard as GSM. The network checks this number during a call. If this number does not match with the list of authorized equipments that the EIR contains, the access is denied.

8.2 GSM AIR INTERFACE

The most important central interface is the air interface in every mobile system. The importance of this interface arises from the fact that it is the only interface the mobile subscriber is exposed to, and the quality of this interface is crucial for the success of the mobile network. The quality of this interface primarily depends upon the efficient usage of the frequency spectrum that is assigned to it. The available and suitable spectrum of frequencies for GSM is limited. The available electromagnetic spectrum has been split into a number of bands by both national and international regulatory bodies. Fortunately, for GSM, 900 MHz and 1800 MHz frequency bands are internationally recognized which brought in large economies of scale, reducing the price of handsets and enabling GSM to flourish.

8.3 GSM MULTIPLE ACCESS SCHEME

The limited radio spectrum is to be shared by all users in Public Land Mobile Telephone Networks (PLMN). For spectral efficiency, GSM works on a combination of frequency division multiplexing (FDM) and the time division multiplexing (TDM) schemes in addition with different interference reduction techniques, i.e., emission power control and optimized handover decision methods.

FDMA In the FDMA system, one specific frequency is allocated to one user engaged in a call. In Europe, GSM 900 band reserves two frequency bands for uplink (890–915 MHz) and downlink (935–960 MHz) so that there exists a duplex distance of 45 MHz. They are divided by frequency into 124 carriers, each separated by 200 kHz.

The Digital Cellular System (DCS) 1800 standard developed by ETSI is based on the GSM recommendations. The frequency ranges for uplink and downlink for this are 1710–1785 MHz and 1805–1880 MHz respectively. A total of 374 carrier frequencies are available to the system. One or more carrier frequency may be

assigned to the base station (BS). Depending on the cell size, propagation environments and cell organization, BS may select the reuse frequency considering the interference effects also.

TDMA Each of the carrier frequencies is divided into eight time slots that form a TDMA frame. A time slot corresponding to an assigned frequency characterizes any physical channel. This is explained in Fig. 8.6.

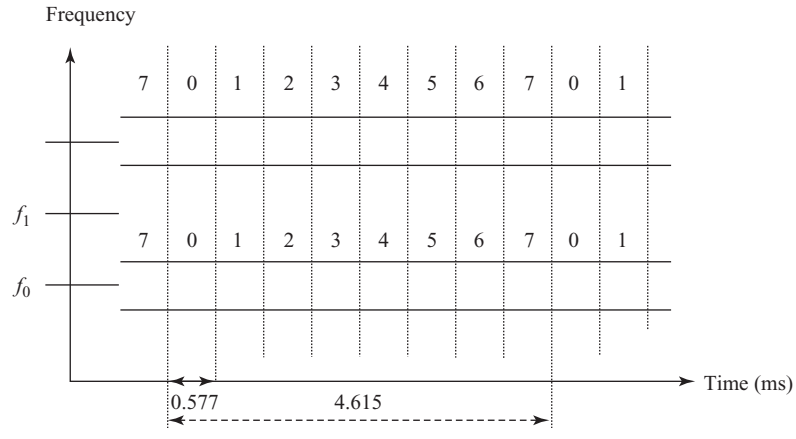


Fig. 8.6 *Physical channel using FDM and TDM accesses*

When there are numerous calls, the network tends to get overloaded, leading to failure of the system. In a full-rate (FR) system, eight time slots (TS) are mapped on every frequency, while in the half-rate (HR) system, 16 TSs are on every frequency. In the TDMA system, only impulse-like signals are sent periodically, unlike in FDMA where signals are assigned permanently. Thus, by obtaining the advantages of both the techniques, TDMA allows seven other channels to be served on the same frequency.

For the full rate system, every impulse on frequency is called a burst. Every burst corresponds to a time slot of 0.577 ms. So

1 TDMA frame = 8 bursts = 4.615 ms

Frequency-division multiplexing divides each of the frequency ranges into 125 channels (25 MHz/200 kHz) of 200 kHz bandwidth. One of the channels is used as guard band leaving 124 channels available for both transmission and reception. Thus, in GSM 124 duplex communication channels are produced.

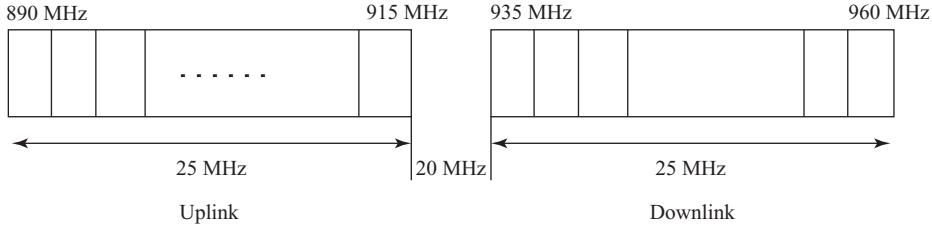


Fig. 8.7 *GSM logical channel*

Figure 8.7 shows the GSM logical channel. There is a gap of 20 MHz between the transmission subbands, i.e., the GSM base station transmit band starts at 890 + 45 MHz (downlink). The mobile device transmits on the lower frequency since at lower electromagnetic spectrum there will be less attenuation. As the base transmission system can radiate larger power, the greater attenuation can be compensated in the downlink and will not be a major problem. A matching pair of one uplink and one downlink in the GSM frequency channel is controlled by a device called transceiver (TRX). All GSM operators within a country using the GSM900

band share 124 channels. Each operator has to buy a license allocated for the frequency and is controlled by the national telecommunication regulator. If there are four operators in a country, then 124/4 = 31 channels are allocated for each of the operator.

Let us explain the operations for frequency allocation with four operators.

For operator 1 The uplink channel's frequency allocation starts at 890 MHz, the next channel frequency is 200 kHz apart, i.e., 890.2 MHz, and so on. So the uplink channels are extended to 896.0 MHz as shown in Fig. 8.8.

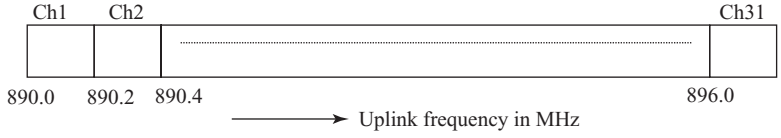


Fig. 8.8 *GSM uplink frequency allocation for operator 1*

Similarly, downlink transmission starts at 935.0 MHz and is extended up to 941.0 MHz for 31 channels as shown in Fig. 8.9.

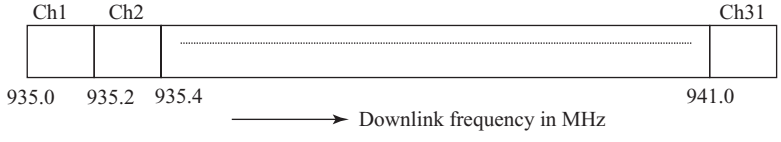


Fig. 8.9 *GSM downlink frequency allocation for operator 1*

The assigned spectrum of 200 kHz per channel is segmented in time by using the TDMA scheme. The time axis is divided into eight time slots, each of length 0.577 ms. Slots numbered from time slots 0 to 7 form a frame with a length of 4.615 ms as shown in Fig. 8.6. The occurrence of one particular time slot in each frame makes up one physical channel. The TDMA scheme uses a gross bit rate of about 270 kbps (with a Gaussian minimum shift-keying modulation, GMSK) and requires sophisticated adaptive receiver techniques to cope with the transmission problems caused by multipath fading. A single time slot in GSM is called a *burst*. Let the channel bit rate = D per unit time. As one TDMA frame has eight bursts, hence each burst has a bit rate $d = D/8$. One time slot is reserved for control channel, leaving seven available TRX for the users.

The TDMA factor of 8 in combination with a carrier spacing of 200 kHz would correspond to the earlier analog system using single-channel per-carrier with a 25-kHz carrier spacing.

8.4 GSM CHANNEL ORGANIZATION

GSM defines a variety of traffic and signaling/control channels of different bit rates. These channels are assigned to logical channels derived from multiframe structuring of the basic eight slotted TDMA frames. One time slot of the TDMA frame is a physical channel. On a physical channel, great variety of information and signaling due to special services is transmitted. Different logical channels are mapped onto a physical channel defined by the number and position of their corresponding burst periods. Logical channels can only be deployed in certain combinations and on certain physical channels. For this purpose, two multiframe structures have been defined. The traffic channel multiframe consists of 26 groups of 8 TDM frames whereas the control multiframe consists of 51 groups of frames.

The physical channel of the TDMA frames carry logical channels that transport user data and signaling information. Logical channels are separated into traffic and control channels. Thus, there are traffic channels and signaling channels in the TDMA frame. The traffic channel (TCH) transmits either data or voice signals and can be HR (Half-Rate) or FR (Full-Rate). HR provides a bit rate for coded speech of 6.5 kbps and FR has double capacity, i.e., 13 kbps.

The signaling channel may contain the information such as broadcast channels (e.g., BCCH, SCH, FCH); common control channels (e.g., AGCH, PCH, RACH), dedicated channels (DCCH, SDCCH) and associated channels (FACCH, SACCH, SDCCH). Each will be discussed separately.

8.4.1 Traffic Channel Multiframe

A Traffic Channel (TCH) is used to carry speech or circuit-switched data traffic. The 26 multiframe is used to define TCH, and their slow and fast associated control channels (SACCH and FACCH) that carry link control information between the mobile and the base stations. The length of a 26-frame is 120 ms. The TCH have been defined to provide six different forms of services, that is, full rate speech or data channels supporting effective bit rates of 13 kbps (for speech), 2.4, 4.8, and 9.6 kbps; and the half-rate channels with effective bit-rates of 6.5 (for speech) kbps, 2.4 kbps, 4.8 kbps for data.

Time slots 12 and 25 are used for SACCH. There is only one SACCH channel transmitted per multiframe as the SACCH may alternate between 12 and 25 time slots on different multiframes. When the mobile device is not transmitting or receiving on its dedicated traffic channel on a frame, it constantly monitors the signal strength received from the cell it is attached to or to the other six neighboring cells. SACCH is used to send the results of this scanning to the network. The length of a SACCH is 456 bits (4 burst × 114 bits per burst). So the reporting time is 480 ms. The mobile device uses this information to increase or decrease its power levels in every 60 ms. Apart from the power control mechanism, SACCH may be used for sending SMS (Short Message Service) to and from the mobile device while the call is in progress.

SACCH is implemented on frame 12 (numbered from 0), providing eight SACCH channels, one dedicated to each of the eight TCH channels. Frame 25 in the multiframe is currently idle and is reserved to implement the additional eight SACCH required when half-rate speech channels become a reality. A traffic channel is only assigned when the mobile device is in dedicated mode, whereas in idle mode the mobile device does not have a traffic channel assigned to it.

The FACCH is obtained on demand by stealing from the TCH, and is used by either end for signaling the transfer characteristics of the physical path, or other purposes such as connection handover control messages. The stealing of a TCH slot for FACCH signaling is indicated through a flag within the TCH slot. Each TDM burst is of 148 bits. The first 3 bits and last 3 bits used for framing are always zero. The stealing F1 bit indicates whether burst contains user data or control information. During the ongoing call process, if the network needs to handle handoff for another call quickly then the network uses the dedicated traffic channel. So, it is referred as Fast Associated Control Traffic Channel or FACCH. Stealing bits indicates to the user that the burst is a control message rather than user data. In the uplink, the mobile terminal indicates the control information in the same way. There is a training field to be used to synchronize the transmitter and receiver as shown in Fig. 8.10.

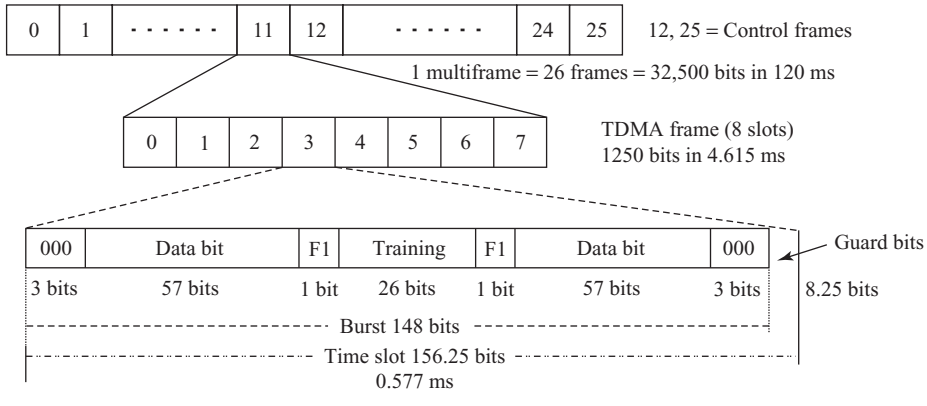


Fig. 8.10 *GSM frame structure and normal burst*

In GSM, a normal burst is 148 bits long. An unused 8.25-bit guard band is used at the end of each burst. Out of 148 bits per time slot, 114 bits are the data bits transmitted in two 57-bit sequences at the beginning and end of the burst. The middle 26 bits are the training sequences, which are used in adaptive equalizing in the mobile and base station receiver in order to analyze the radio channel characteristics before the user data is decoded. A GSM subscriber uses one time slot to transmit, one to receive and the remaining six slots to measure the signal strength on the neighbouring base stations within a frame.

8.4.2 Control (Signaling) Channel Multiframe

The control channel multiframe consists of 51 frames, which incorporate control, timing and signaling. User data is not only the flow of information to be transported during a call, but also the signaling messages that helps the mobile station and network to manage a call. The 51-frame multiframe has a more complex structure. These channels are discussed below.

BCCH Broadcast control channel is used in the BSS to mobile direction to broadcast system information such as the synchronization parameters, available services, and cell ID. This channel is continuously active, with dummy bursts substituted when there is no information to transmit, because mobiles for handover determination monitor its signal strengths. All mobile stations can monitor the strength of this signal to ensure that they are still within the cell.

DCCH Dedicated control channel is divided into three subtypes. This point-to-point control channel provides access to the slow associated control channel, standalone dedicated control channel and fast associated dedicated control channel.

SDCCH Standalone dedicated control channel is used for call set-up and location updating of the mobiles. It is also used for SMS to and from mobiles that are in idle state. Like the TCHs, the SDCCH has its own SACCH (Slow Associated Dedicated Control Channel) and is released once call setup is complete.

CCCH Common control channel is divided into three subchannels, one used for uplink and two used for downlink. It is used for transferring signaling information between all mobiles and the BSS for call origination and call paging functions.

RACH Random access channel is used for uplink communications. It allows a mobile station to send request for time slot on the dedicated control channel that can be used to assign a TCH for voice call. The mobiles use the slotted Aloha scheme over this channel for requesting a DCCH from the system at call initiation.

PCH Paging channel is used to send a paging message to the mobile user during incoming call arrival. Within specific time intervals the MS will listen to the PCH to see if the network wants to communicate with the mobile. The information on the PCH is a paging message including IMSI or a TMSI.

AGCH Access grant channel used by the system to assign resources to a mobile such as a DCCH channel. The base station announces the assigned slot on the AGCH.

AGCH and the PCH are never used by a mobile at the same time, and therefore are implemented on the same logical channel.

FCCH Frequency correction channel is used to ensure that the mobile device tunes its frequency reference to that of the base station it is connected to. FCCH carries information from the BSS for carrier synchronisation.

SCH Synchronization channel is used to frame synchronize the mobile device by broadcasting the base station identity code (BSIC) on SCH so that the mobile device understands the correct base station of the network it tries. This BSIC can only be decoded if the BS belongs to a GSM network.

All the control-signaling channels, except the SDCCH, are implemented on the time slot 0 in different frames of the 51 multiframes using a dedicated RF carrier frequency assigned on a per cell basis.

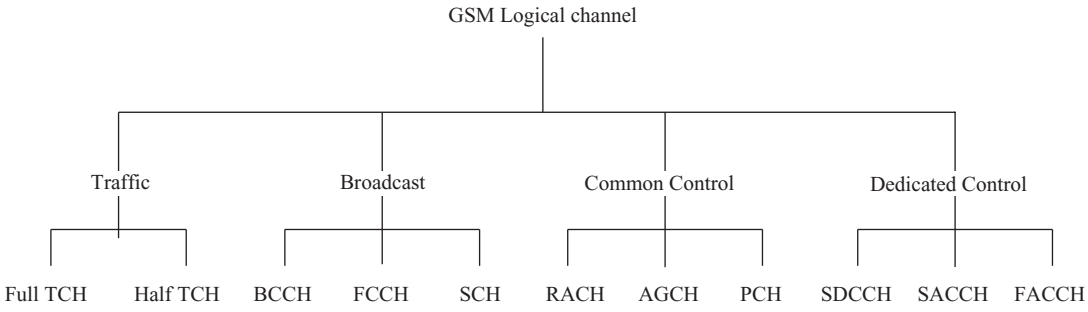


Fig. 8.11 *Divisions of GSM logical channels*

8.4.3 Frames, Multi-frames, Super-frames and Hyper-frames

TDMA frames are numbered which are used as input parameters for the encryption process. For hyper-frames, this number becomes large. A frame consists of 8 time slots of 4.615-ms duration. A multi-frame is a block of 26 frames used to transfer information having a total duration of 120 ms. A super-frame consists of 26×51 TDMA frames with a duration of approximately 6.12 seconds. Since 26 is not a factor of 51, these frames slide across each other, so that at the end of the 26×51 period each of the 26 frames has aligned once with every one of the 51 control frames. Sliding is required to allow the mobile device to check the BCCH of the neighboring cells for handover process. A hyper-frame consists of $2048 \times 26 \times 51$ TDMA frames lasting over a time of 3 hours, 28 minutes, 53 seconds and 760 ms. Figure 8.12 explains the frame processes of GSM.

When the signaling frame is described, it is important to know precisely which frame is transmitted. This is required to know the exact SDCCH number to be received. There is no other way to know the SDCCH numbers. To avoid ambiguity, the frames are numbered specially. There are counters T1, T2 and T3 for super-frames, voice frames, i.e., 26 multi-frames and 51 signaling frames respectively.

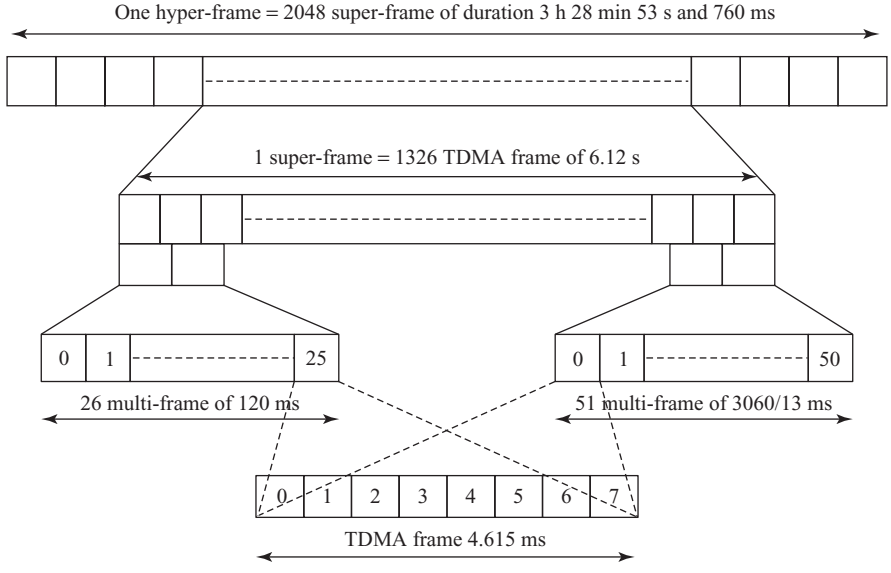


Fig. 8.12 *GSM multi-frame structures*

Whenever a super frame is completed, T1 is incremented by 1. It has values in between 0 to 2047. T2 varies from 0 to 25 and T3 from 0 to 50. The counter initialises to zero and then transmission of frames

begins. On completion of voice or signaling frames, the respective T2 and T3 counters are set to zero and start counting again and on completion of transmission of super-frame, T1 is incremented by 1. T1 restarts counting after reaching the count 2,047, which needs more than 3 hours. Knowing the values of T1, T2 and T3 and also the types of multi-frames assigned to each of the available TDMA time windows, it is possible to know what is being displayed in each time window at the instant considered. The frame identification numbers consisting of the values of T2 and T3 are transmitted through SCH channel. Getting these numbers, the mobile station can look for BCCH and information about the system. This information is important for a mobile to know exactly how long it has to wait for some data and when to start transmission.

Different channel coding methods [11, 12] are applied on the various channels for different services. The channels have Forward Error Correction (FEC) protection through convolutional coding. Depending on the channel types, coded data are encapsulated into the related bursts which are transmitted with a gross bit rate (modulated bit rate). The GSM system uses a digital 0.3 GMSK (Gaussian Minimum Shift Keying) [13] modulation format. Minimum shift means there is no discontinuity between the phase of the carrier and the time.

8.5 GSM CALL SET-UP PROCEDURE

When the mobile user powers up the mobile terminal, it tries to register first with a network. The home network of the subscriber is stored in the SIM card and is first checked for authenticity of the user. On availability of the home network, the mobile device sends a connection request, otherwise it tries to connect to the last network before it was switched off. This information is stored in the SIM module. Arriving at a new place, a mobile device searches all the frequencies within the band to establish connection to the existing network. This means the mobile device looks for strong BCCH signal that includes FCCH and SCH. FCCH tries to synchronize the mobile device with the BS frequency by emitting a sine wave as to whether SCH contains the BS identity and frame numbers. BCCH also provides the location area information of the BS to the mobile device. The connection request by the mobile is sent through RACCH to which the BS listens continuously in order to locate mobiles wanting to register themselves. RACCH works on the slotted Aloha protocol to avoid contention from the different mobile users requests at the same time. The mobile user waits for a random amount of time for registration.

BSs regularly broadcast their presence, while the mobile device continuously checks the paging channel and replies on the RACCH (Random Access Control Channel) dedicated channel. The acceptance signal by the network is sent on the Access Grant Channel (AGCH). SDCCH is used by a mobile device to continue its negotiation at a lower bit rate than the TCH so as to transfer data for signaling purposes.

8.5.1 Authenticity Check-up

The IMSI is sent by the mobile to the MSC to be processed with the help of HLR/AuC of the subscriber home network for authenticity. On reply, the MSC sends the random number (RAND), a key (K_c) and a result SRES (Signed Response). The SIM card of the mobile device is required to produce another SRES′ that is to be sent back to the MSC and then compared to the result from the AuC using this RAND. Obviously, an invalid user will reply with a wrong result. K_c is used for data encryption between the MS and BTS. After the completion of authentication, the MSC requests the IMEI of the mobile device. On receiving the data, it is checked with the EIR to see whether the device is not a stolen one. If IMSI and IMEI are checked successfully, the MSC sends a request to HLR for information about the subscribers, service types and other details. The MSC then registers the mobile device with the current VLR and lets the HLR know the current location of the device. Within the VLR, the mobile device uses a temporary identity called TMSI, which is obtained from MSC and ensures better security of the user in the visited network and does not expose the actual identity of the IMSI. At this point, the initial signaling procedure is complete and the mobile device is assigned an SDCCH or a TCH to proceed the call. Figure 8.13 depicts the call set-up procedure.

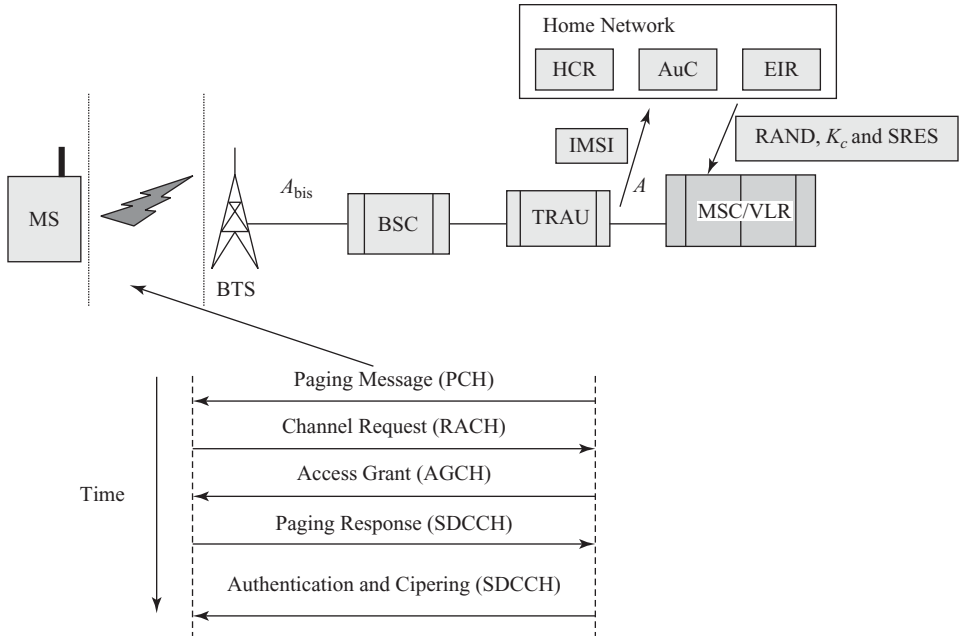


Fig. 8.13 *GSM call set-up procedure*

8.6 GSM PROTOCOLS AND SIGNALING

The signaling protocol in GSM is structured into three general layers depending on the interface. Layer 1 is the physical layer, which uses the channel structures discussed above over the air interface. Layer 2 is the data link layer. Across the U_m interface, the data link layer is a modified version of the Link Access Protocol D (LAPD) used in ISDN, called the LAPDm. Across the A interface, the Message Transfer Part Layer 2 of Signaling System Number 7 (SS7) is used.

The GSM air interface consists of TDMA/FDMA time slots and frequency bands. LAPDm is used over the air interface between the Base Station Transceiver (BTS -TRX) and the mobile device. To transport information to a desired destination, some additional control information is needed apart from the actual data. This is known as *signaling message*. Basically, the signaling channels are logically multiplexed on an aggregate of the TDM slots.

Layer 3 of the GSM signaling protocol is itself divided into 3 sublayers:

1. Mobility Management (MM),
2. Radio Resource Management (RRM), and
3. Connection Management (CM) for calls routing.

Unlike the normal Layer 3 functionalities in the OSI model, the GSM Layer 3 protocol is somewhat different. It is used for the communication of network resources, mobility, code format and call-related management messages between the different network entities. RRM is implemented between the MS and BSS, whereas MM and CM is the communication between the MS and MSC. GSM Layer 3 can also be implemented for Message Transport Part (MTP) protocol and the signaling connection control part of the CCITT (Comitte Consultatif International Telegraphique et Telephonique) SS7 signaling that works between the BSS and MSC over the A interface. This is related to transport and address functions for signaling messages for call-routing procedures via the MSC. It is to be noted that Layer 3 protocols in GSM provide functionalities similar to the transport, session and presentation layers in the OSI model.

8.6.1 A_{bis} Interface

As shown in Fig. 8.13, the A_{bis} interface is the interface between the BTS and the BSC. It is a PCM interface, i.e., it is defined by the 2 Mbps PCM link. Thus, it has a transmission rate of 2.048 Mbps, having 32 channels of 64 kbps each. As the traffic channel is 13 kbps on the air interface, while A_{bis} is 64 kbps, hence multiplexing and transcoding conversions are done from 64 kbps on the A_{bis} interface to 13 kbps on the air interface.

8.6.2 *A* Interface

Between the TRAU and MSC or physically between the MSC and BSC, there is the *A* interface. This interface consists of one or more PCM links, each having a capacity of 2048 Mbps. There are two parts of the *A*-interface—one from the BTS to the TRAU where the transmitted payload is compressed, and one between the TRAU and MSC where all the data is uncompressed. The TRAU is typically located between the MSC and BSC and should be taken into consideration when dealing with this interface. SS7 is present on the *A* interface.

8.6.3 Link Layer LAPDm Protocol

The data link-layer over the radio link connects the MS to the BSS based on a LAPD-like protocol, called LAPDm. LAPDm does not use any flag for frame delimitation, rather the physical layer defining the frame boundaries does frame delimitation.

There is a length indicator in LAPDm to distinguish the information-carrying field from the fill-in bits used to fill the transmission frame. LAPDm uses a 3-bit Service Access Point Identifier (SAPI) as an address field. SAPI 0 is used for call control, MM and RR signaling. SAPI 3 is used for SMS. All the other fields are reserved for future purpose. The data-link layer also provides the priority assignment as low, normal and high priority to messages that are transferred in dedicated mode on SAPI 0.

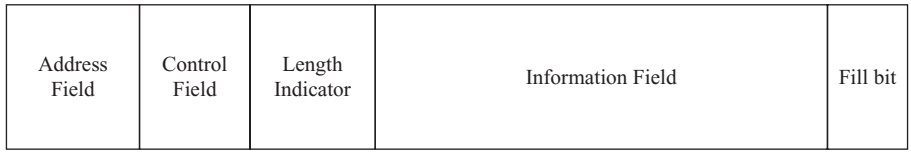


Fig. 8.14 *General frame format for LAPDm*

The general frame format for LAPDm is shown in Figure 8.14. There is the address field that has the frame format as shown in Fig. 8.15. The two-bits Link Protocol Discriminator (LPD) is used to specify a particular recommendation of the use of LAPDm, the C/R is a single bit which specifies a command or response frame as used in LAPD, and a 1-bit Extended Address (EA) is used to extend the address field to more than one octet. The EA bit in the last octet of the address should be set to 1, otherwise it is set to 0. The 8-bit is reserved for future uses.

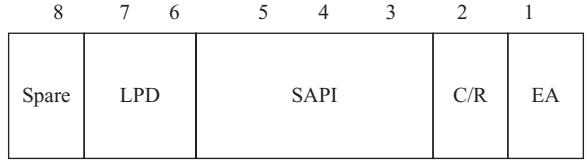


Fig. 8.15 *Address field format for LAPDm*

The control field is to carry sequence numbers and to specify types of the frame. There are three types of LAPDm frames, one for supervisory functions, unnumbered information transfer and control functions for unacknowledged mode and numbered information transfer for multiframe acknowledged mode. LAPDm has no Cyclic Redundancy Check (CRC) flag, as it is done by the combination of block and convolutional coding used in the Layer 1 (physical).

8.6.4 Message Layer Protocols

Figure 8.16 represents the GSM signaling protocols. Priority is generally given to RRM, MM and CM. The RR layer is used to establish, maintain and release RR connection to set up point-to-point communication between the mobile device and the network. This connection is used for data and user signaling. The procedures include the selection, reselection and handoff process and the reception of BCCH and CCCH when RR is established.

Interface A establishes connection between BSS and MSC. MTP level 2 of the SS7 protocol is similar in providing OSI layer 2 functionalities for reliable transport of the signaling messages through error detection and corrections, and retransmission of the signals.

As mentioned, the earlier **RRM sublayer** is used to establish connection between the mobile device and BSS over the radio interface for transmitting call-related signaling and traffic channel for a particular mobile user to the BSS. The RR layer is concerned with the management of an RR-session, which is the time that a mobile is in the dedicated mode.

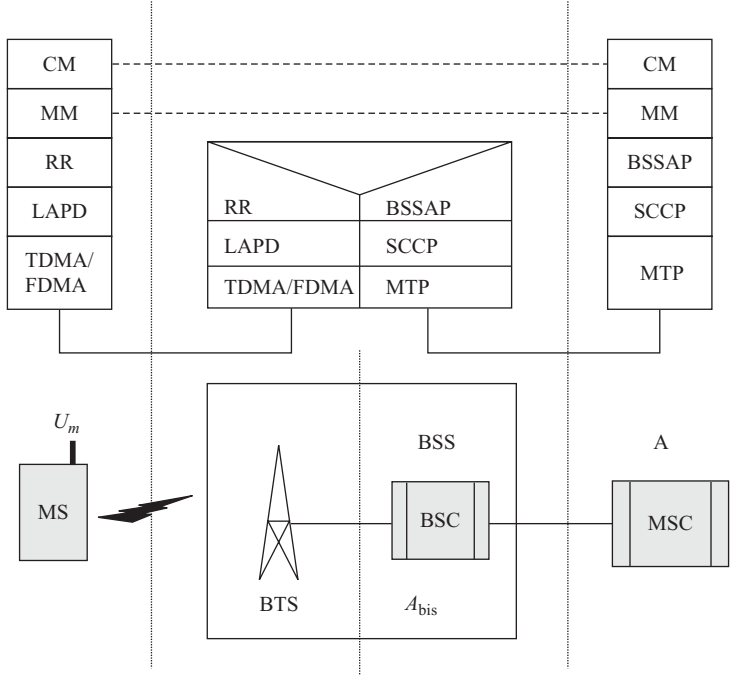


Fig. 8.16 *GSM signaling protocols*

An RR-session is always initiated by an MS through the access procedure, either for an outgoing call, or in response to a paging message. In addition, it handles the management of radio features such as power control, discontinuous transmission and reception, and timing advance.

Handover or Handoff is the process of switching of an on-going call to a different channel or cell. The handover management and execution is one of the important functions of the RR layer. There are four different types of handover in the GSM system, which involve transferring a call between

1. channels (time slots) in the same cell, called *intracell handoff*
2. cells (BTS) under the control of the same BSC, called *intercell handoff*
3. cells under the control of different BSCs, but belonging to the same mobile MSC, called *inter-BSC handoff* or *intra-MSC handoff*
4. cells under the control of different MSCs, called *inter-MSC handoff*

During its idle time slots, the mobile scans the BCCH for up to 16 neighboring cells, and forms a list of the six best candidates for possible hand off, based on the received signal strength. This information is passed to the BSC and MSC periodically (at least once per second), and is used by the hand-off algorithm.

The Mobility Management layer (MM) is built on top of the RR layer, and handles the functions that arise from the mobility of the subscriber, as well as the authentication and security aspects. The MM sublayer is terminated at the MSC. It also provides connection management services to the CM layer. There are three distinct functionalities of the MM sublayer. They are – *MM specific procedures, MM common procedures*, *and the MM connection related procedures*.

MM Connection-Related Procedures are used to establish, maintain and release an MM connection between the MS and the MSC over which an entity of the CM sublayer can exchange information with its peer. The MM connection can serve multiple CM entities. The protocol discriminator and transaction identifier within the related signaling messages exchanged, identify each of those entities. Because of this separate transaction identifier for different calls, an MS can handle parallel calls. The CM sublayer is functionally divided into different entities, such as Call Control (CC), Short Message Service (SMS) support, Supplementary Service Support (SSS) and Location Services Support (LSS).

An MM connection is initiated by a **CM service request** message, which identifies the requesting CM entity and the type of service required for the MM connection. Registration, Security and Connection Management are the services provided by MM. The MM handles authentication request and response messages as well as identity request and response messages.

The **MM specific procedure** consists of location update and IMSI attach procedures, whereas, the **MM common procedures** consist of IMSI detach, TMSI reallocation, and authentication/ identification.

Location Update Procedure Location update is the procedure which keeps track of the mobile user while he/she is roaming and is always initiated by the MS. Location management is concerned with the procedures that enable the system to know the current location of a powered-on mobile station so that incoming call routing can be completed. A powered-on mobile is informed of an incoming call by a paging message sent over the PAGCH channel of a cell. If the network sends the paging message to every cell (called blanket paging), there is an unnecessary wastage of radio bandwidth. On the other hand, if every cell generates the location update message during movement then it requires a large location update message, though it requires only single cell paging. Considering trading off the two, GSM uses the concepts of **Location Area (LA),** group of cells within which a paging message is to be sent. MSC and two location registers, the **Home Location Register (HLR) and the Visitor Location Register (VLR),** take an active role in the LA update process. During the power-on mode of the mobile device or any movement to a new position; it must register with the network to indicate its current location. To do this, a location update message with the location information is sent to the new MSC/VLR and then this location information is sent to the subscriber's HLR. The information sent to the HLR is normally the SS7 address of the new VLR (a routing number). For a valid user, the HLR sends a subset of the subscriber information, needed for call control, to the new MSC/VLR, and sends a message to the old MSC/VLR to cancel the old registration.

The network periodically broadcasts the availability of the BS; the MS monitors this location information regularly and compares it with the previous information stored in the memory. The MS can also receive an indication from the network that is not known by the present VLR while trying to establish an MM connection. In this situation, the MS detects the new location and the process of location update is invoked. A procedure related to location updating is the **IMSI attach and detach**. A detach lets the network know that the mobile station is unreachable, and there is no need to allocate channels and send paging messages. An attach is similar to a location update, and informs the system that the mobile is reachable again. The activation of IMSI attach/detach is up to the operator on an individual cell basis.

Whenever location update for a mobile is required, the network sends a TMSI in ciphered mode that is stored in MS. This number is used for mobile identification in the visited location. In combination with

TMSI, the location area identifier (LAI) is used for identification of the unambiguous mobile outside the area where it is assigned. The IMSI attach procedure is then performed only if the stored location area at the time is the same as the one being broadcast on the BCCH channel of the serving cell. Otherwise, a normal location update procedure is invoked regardless of whether the network supports IMSI attach/detach procedures.

IMSI Attach and Detach Procedures These are network initiated and indicated through a flag in the system information that is broadcast on the BCCH. From the IMSI attach or detach procedure, the MS can mark whether it is attached/detached to the VLR (or HLR) during the power-up or power-down phases of the MS, or on removal of SIM card. Clearly, the IMSI detach procedure disables the location update function. The IMSI detach procedure is invoked by the MS to indicate its inactive status to the network. No response or acknowledgement is returned to the MS by the network on setting the active flag for the IMSI. The IMSI detach procedure is not started if at the time an **MM-specific procedure is active**. In that situation, the IMSI detach procedure is delayed until the MM-specific procedure is finished, otherwise the IMSI detach request is omitted. If a radio connection between the MS and network exists at the time of a detach request, the **MM sublayer will release** any ongoing MM connections before the MM detach indication message is sent.

Another process supported in the GSM MM common procedure is called **TMSI reallocation**. This is performed when MSC changes the coverage area for the mobile user. The TMSI reallocation provides identity confidentiality of the user from any intruder. If the network is unaware of the TMSI of the mobile then the MS has to send the IMSI value to the network on request. In this case, the identification procedure has to be performed before the TMSI procedure can be initiated.

8.7 AUTHENTICATIONS AND SECURITY

Authentication is the process to prove the identity of valid users claiming services of the network. The wireless radio is exposed to anyone for accessing. So, it is very important for any mobile network to identify its users. Two main entities are involved—the SIM card in the mobile device and the authentication center, AuC. A secret key is provided to each user that is stored both in the SIM card and the AuC. A signed response (SRES) is generated randomly using this secret key and the ciphering algorithm, A5. This is sent back to the AuC to be compared with the number generated by the AuC. If matched, the subscriber is an authentic one. The authentication procedure is always initiated and controlled by the network.

Another level of security is performed on the mobile equipment itself. Each of the GSM terminal is identified by a unique International Mobile Equipment Identity (IMEI) number. A list of IMEIs in the network is stored in the Equipment Identity Register (EIR). For identification, the network may request a mobile user to provide the IMEI or IMSI. In reply, the MS should send its identity while RR connection exists between the mobile and the network. The status returned in response to an IMEI query to the EIR is one of the following:

White-listed The terminal is allowed to connect to the network.

Grey-listed The terminal is under observation from the network for possible problems.

Black-listed The terminal has either been reported stolen, or is not type approved, i.e., it is not the correct type of terminal for a GSM network. The terminal is not allowed to connect to the network then.

8.8 GSM AND SIGNALING SYSTEM 7

Signaling between the different entities in the fixed part of the network, such as between the HLR and VLR, is accomplished through the **Message Application Part (MAP)**. MAP is built on top of the **Transaction Capabilities Application Part (TCAP)**, the top layer of the **Signaling System Number 7**. SS7 or Common Channel Signaling System 7 (CCS7) is a telecommunications signaling system standard given by the International Telecommunication Union (ITU-T) under the Q.700 series of recommendations. SS7 was to provide signaling system to be used globally. A majority of the fixed line telephone networks are based on

SS7 signaling. This is why SS7 is also used in cellular systems, as it allows the simple integration of cellular and fixed line networks. A typical SS7 network consists of three types of devices:

1. *SSP*—Service Switching Points
2. *STP*—Service Transfer Points
3. *SCP*—Service Control Points

These three together are referred as SS7 nodes. Although logically they are separate entities, for a single device, they perform a number of functions.

The SSP is responsible for originating, terminating and routing the call to correct destination. The call request is sent to a switching center that can handle a large number of calls simultaneously. In a GSM, the MSC is considered as an SSP.

STP routes the call from source to destination and also finds an alternative path in case of network failure.

SCP is required to process toll-free numbers by searching the telephone numbers in a database. A similar kind of thing happens in cellular networks. The network needs to query the HLR of the mobile subscriber with the subscriber MSISDN. In a GSM network register, the HLR, VLR, EIR and AuC are of SCP type.

8.8.1 Protocol Stack for SS7

The standard protocol stack for SS7 is given in Figure 8.17 in respect of OSI seven layer systems.

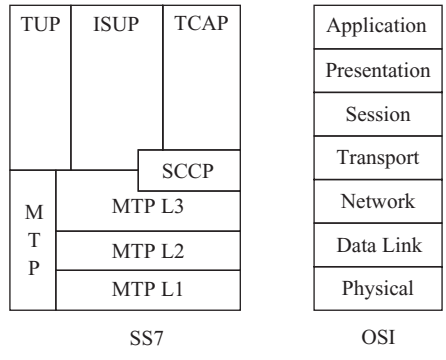


Fig. 8.17 *Signaling protocol stack for SS7*

The lower three parts of the SS7 stack are collectively known as Message Transfer Part (MTP).

MTP Level 1 (MTP L1) is the physical layer of the OSI model. Messages are carried over 56 or 64 kbps TDM links for narrowband services. However, high-speed channels of 32×64 kbps or 24×64 kbps are also supported.

MTP Level 2 (MTP L2) ensures a reliable connection between two data-link network elements. It has error checking, flow control and sequencing mechanism of the data transmission. In case of error, MTP L2 asks for retransmission. This part is equivalent to the data-link layer of OSI model.

MTP Level 3 (MTP L3) is the extension of Level 2 for the transportation of signaling messages over the network. It deals with the routing and re-routing of the signal, similar to IP layer functionalities of the OSI model.

The Signaling Connection Control Protocol (SCCP) and the SS7 MTP protocols are together used to implement the data link and Layer 3 transport functionalities for carrying the call control and MM signaling messages over the BSS–MSC link. SCCP sits on top of MTP and provides both connectionless and connection-oriented services. It is similar to the TCP layer of Internet suit protocols.

The MM and CM sublayer signaling information from the MS is routed over signaling channels, like DCCH, SACCH, and FACCH, to the BSS and are transparently relayed through the Direct Transfer Access Process (DTAP) to an SCCP, which is of the CCITT SS7 type of logical channel assigned for that call on the BSS–MSC link to the peer CC entity for processing in MSC. For the reverse transmission from MSC to BSS, the SCCP connection is relayed to the DTAP process on the BSS. Using the LAPDm data link protocol, the BSS sends this call to the MS. Within the information field of SCCP, there is a distribution data unit that provides a link between the Layer 2 protocols and SS7 on the BSS–MSC link. The parameters for distribution data units are the discriminator and the Data Link Connection Identifier (DLCI) parameters. Again, there are two subparameters of the DLCI, one is used to identify radio channel types, whether DCCH, SACCH or FACCH, and the other to identify the SAPI value in the LAPDm protocol. Basically, SCCP multiplexes the different calls onto the same physical channel on the BSS–MSC link. Information flows to the associated SCCP connection that is established on the BSS–MSC link related to a specific call supported by a BSS. The SCCP also provides address translation capabilities, known as Global Title Translation (GTT).

Telephone User Part (TUP) is designed for traditional analog circuits to set-up calls and release calls. TUP was the first application defined and does not provide ISDN services.

ISDN User Part (ISUP) is defined for signaling messages used to set-up calls, modification and tear down of calls for ISDN services. ISUP supports basic telephony in a similar manner to TUP. Some of the example of ISUP messages are answering the call, charging information, connecting, identification request, response and also the release.

8.8.2 Transaction Capabilities Application Part

TCAP defines the information flow and can also report results. Queries and responses between SSPs and SCPs are sent via TCAP. Registration of roaming users is done through TCAP. When the mobile subscriber roams into the new MSC coverage, the VLR requests information about the subscriber from the HLR. The HLR uses Message Application Part (MAP) protocol to carry this information within TCAP messages. Although the SCCP can locate the database, the actual query for the data is performed by a TCAP message. Apart from this, the TCAP also transports the billing information.

There are two other important signaling protocols in the GSM user part that help in tracking the roaming user from one location to another within the network. These are MAP and BSSAP (Base System Application Part). MAP is the key protocol for cellular communication networks to be accessed for roaming information, paging, managing handover and sending SMS messages. MAP messages are carried by TCAP.

BSSAP messages work between the MS and MSC and also between the BSC and MSC. It is divided into two parts.

The Direct Transfer Protocol Part (DTAP) is used to transfer the message between the MS and MSC. These messages which include MM and CM are not interpreted by the BSS, but are transparent.

The BSSMAP is the process within the BSS that controls radio resources under the instruction of MSC. It is used to implement all procedures between the MSC and the BSS that require interpretation and the processing of information related to single calls, and resource management. It supports procedures like resource management, handover management and paging of the mobile device.

8.9 ROUTING OF A CALL TO A MOBILE SUBSCRIBER

A call may originate from a mobile to another mobile, or from a fixed landline telephone to a mobile or the reverse. For routing a call to a mobile user, the first network needs to find the location of the mobile.

When a landline user calls a mobile subscriber, the PSTN will use the Mobile Station ISDN number (MSISDN) instead of using IMSI. MSISDN is the dialed directory number used to reach a mobile

subscriber. This number includes a country code and a National Destination Code that identifies the subscriber's operator. The first few digits of the remaining subscriber number are used to identify the subscriber's HLR within the home PLMN.

For a mobile terminated call, this number is routed to the Gateway MSC (GMSC) in the home network of the mobile subscriber. GMSC sends a MAP request to the HLR with this MSISDN number. Basically, the GMSC is a switch. It interrogates the subscriber's HLR to obtain routing information that uses a table linking the MSISDN to their corresponding HLR. In reply for routing information, GMSC gets the Mobile Station Roaming Number (MSRN), which is defined in the E.164 numbering plan. This MSRN is simply related to the geographical numbering plan and is invisible to subscribers. HLR for the routing information is required to extend the call to the visiting MSC of the mobile at the time. The visiting MSC or the VLR within the visiting MSC is identified by the MSRN allocated by the VLR. This MSRN is sent to the HLR during the location update process and back to GMSC, which then routes the call to the new MSC. At the new MSC, the IMSI corresponding to the MSRN is looked up, and the mobile is paged with a paging broadcast to all BSSs in its current location area. After paging response, the exact BSS is located and then the MM and RR connections are established. At this time, both the authentication and cipher mode setting are performed.

The VLR sends the required parameters to the MSC for call set-up and a new TMSI is assigned to the MS for that call. The MSC sends call set-up messages to the MS. In reply, the MS sends call confirmed messages including bearer capability of the MS. The BSS at this stage assigns a traffic channel to the MS for the call. Receiving the connect message from the MS to the calling subscriber, the network sends an acknowledgement to the MS that enters into the active state.

It is to be mentioned here that when an SMS is sent to a mobile device, it arrives at a service center which is responsible for relaying the message in a store-forward technique. The service center identifies the GMSC for the SMS for this particular mobile and forwards the message.

Summary

In this chapter, an overview of GSM architecture, protocols and signaling is provided. A thorough reading of this chapter gives an understanding about the working of GSM networks. Though the standards and details of GSM systems may cover more than 5000 pages, however, the basic principle of GSM operation, various protocols involved and calling procedures using this system is given briefly. For details of GSM network systems, readers are advised to read the references provided.

The GSM systems operating at 900 MHz and 1800 MHz are the first successful techniques for cellular communication. The novel idea of using a SIM card in the system provides personal mobility and the user can roam globally. The evolution path of GSM systems from 2G to 3G UMTS can fulfill the user's requirements for high-speed data and present-day multimedia services.

References

[1] Audestad, Jan A., *Network Aspects of the GSM System, EUROCON 88*, June 1988.

[2] Balston, D.M., and R.C.V. Macarios, *The Pan-European system: GSM*, editors, *Cellular Radio Systems*. Artech House, Boston, 1995.

[3] Rahnema, Moe, *Overview of the GSM System and Protocol Architecture, IEEE Communications Magazine*, April 1995.

[4] Southcott, C.B. *et al*., *Voice Control of the Pan-European Digital Mobile Radio System, IEEE GLOBECOM '89*, November 1989.

[5] Vary, P. *et al*., *Speech Codec for the European Mobile Radio System*, *IEEE GLOBECOM '89*, November 1989.

[6] Mallinder, J.T. Bernard, *Specification Methodology Applied to the GSM System, EUROCON '88*, June 1988.

[7] Scourias, John, *Overview of the Global System for Mobile Communications*, jscouria@www.shoshin.uwaterloo.ca
[8] Stuckmann, Peter, *The GSM Evolution—Mobile Packet Data Services*, John Wiley & Sons, Ltd., First Edition, 2005.
[9] Bannister, J., P. Mathur, and S. Coope, *Convergence Technologies for 3G Networks*, IP, UMTS, EGPRS and ATM, John Wiley and Sons Ltd, England, 2004.
[10] Garg, V.K., and J.E.Wilkes, *Principles and Applications of GSM*, Prentice Hall, NJ, 1999.
[11] Steedman R., *Speech Codecs for Personal Communications*, IEEE Communications Magazine, pp. 76–83, Nov 1993.
[12] Steele, R., *ed Mobile Radio Communications*, IEEE Press, 1994.
[13] Murota, K., and K. Hirade, *GMSK Modulation for Digital Mobile Radio Telephone*, IEEE Trans. on Communications, COM–29, pp. 1044–1050, July 1981.

Questions for Self-Test

8.1 Digital transmission gives high speech quality and increases radio spectrum efficiency by the use of the following multiple access technology.
a. FDMA b. TDMA c. CDMA d. All the above

8.2 The basic frequency regions for GSM is
a. 900 MHz b. 1800 MHz c. 1900 MHz d. All the above

8.3 The VLR stores information about subscribers visiting a particular area within the control of a specific MSC.
a. True b. False

8.4 Personal mobility can be supported because of
a. SIM b. HLR c. VLR

8.5 IMSI, a unique identity number, resides in
a. SIM b. HLR b. EIR

8.6 In GSM, the BTS can handle
a. seven users at a time b. eight users at a time

8.7 One time slot of TDMA frame is
a. logical channel b. physical channel

8.8 The traffic channel multiframe structure consists of
a. 26 groups of 8 TDM frame b. 51 groups of frames

8.9 For signaling frame, the exact SDCCH number to be received is required to know about the transmitted frame.
a. True b. False

8.10 There are 4 super-frames, 26 multi-frames and 51 signaling frames to avoid ambiguity.
a. True b. False

8.11 VLR provides the proxy services of an HLR within the control of a specific MSC.
a. False b. True

8.12 BSC communicates to BTS through TDM over $A_{bis.}$
a. True b. False

8.13 IMEI is checked during a call handling.
a. True b. False

8.14 DCS—Digital Cellular System is based on GSM recommendation.
a. Yes b. No

8.15 The major function of MSCs are
a. Mobility handling b. Switching c. Handoff
d. Call handling e. All of these

8.16 A copy of the user's secret key is kept in the
a. AuC b. EIR

8.17 The role of TRAU is to convert 16 kbps speech to 64 kbps rates over PSTN or ISDN.
a. True b. False

8.18 Describe the major functionalities of MSC. Why is GMSC needed in GSM architecture?

8.19 What is the importance of TMSI?

8.20 What are the frequency spectrums for DCS system?

8.21 At what frequency does GSM base station start transmitting?

8.22 Why does a mobile station transmit at lower frequency?

8.23 Explain the GSM uplink and downlink operation with pictorial representation.

8.24 If the channel bit rate for GSM is 13 kbps, what will be the bit rate for each burst?

8.25 If a GSM multi-frame consists of 26 groups of 8 TDM frames, what will be the time duration for each multiframe?

8.26 What is the length of SACCH in bits and time duration?

8.27 Draw the GSM network architecture. What are the three main parts of GSM networks? Describe the functionality of each part.

8.28 Describe the GSM TDMA/FDMA access technology.

8.29 With the help of pictorial representation, describe the different GSM logical channel structures highlighting their functionalities.

8.30 Describe the importance of GSM frame structures.

8.31 Describe the function of SACCH and FACCH.

8.32 How are authentication and security maintained in GSM networks?

8.33 State the importance of the following numbers in GSM networks:
IMSI, EIR, TMSI, MSISDN

8.34 What are IMSI attach and detach procedures?

8.35 How is a call routed to a GSM mobile?

8.36 Discuss the role of the following entities with respect to GSM network architecture: Mobile switching center (MSC), Home Location Register (HLR) and Authentication Centre (AuC).

8.37 Explain the operation of multiple access and duplexing technique for GSM.

8.38 What are the Layer 1 and Layer 2 protocols for GSM?

8.39 What are the three main parts of GSM Layer 3 protocols? Describe their main functionalities.

8.40 Draw the signaling protocol stack for GSM with respect to the OSI model. Describe the main functions of each layer.

8.41 Why is a guard band used in GSM frame? What is the importance of 26 training bits? Why are they placed at the middle?

2.5G Networks—The General Packet Radio Services: GPRS

9

Introduction

Global system for mobile communications (GSM) is mainly suited for circuit-switched voice communications. With the increasing demand for data services for a mobile subscriber, the mobile penetration rate is increasing all over the world as does the craze for value-added data services like email, Internet access, sending and receiving images.

The General Packet Radio Services (GPRS) is the GSM evolved mobile network that allows data services by sending and receiving IPv4/IPv6 data packets over the mobile networks. On the evolutional path of mobile wireless networks, GPRS stands for 2.5G. It is of the 'always-connected' type in contrast to a GSM call or a fixed-line call where every time a new connection is made, there is a delay. The European Telecommunications Standards Institute (ETSI) originally specified GPRS, and then it was moved under the project of 3GPP in the year 2000.

Within the next few decades, it is expected that there will be an extensive demand for wireless data services; and mainly high-performance wireless Internet access will be the requested application domain. Existing cellular data services do not fulfill the needs of users and providers, as from the user's perspective, data rates are too slow, the connection set-up takes too long and is rather complicated. The service is expensive too for most users. From the technological aspect, the drawback results from the fact that the current wireless data services are based on circuit-switched radio transmission. A complete traffic channel is allocated per user basis at the air interface for the entire call duration. During busy traffic hours, this will create inefficient radio resource utilization. It is obvious that for bursty traffic, packet-switched bearer services result in a much better utilization of the traffic channel.

The major advantages of GPRS arise from the use of packet switching. Data can be exchanged directly in the form of a packet to the Internet or other data networks. Users can share the time slots rather reserving continuously, thus increasing spectrum efficiency. GPRS uses the same frequency standards allocated for GSM. The major concern for GPRS was to enable mobile devices for handling data traffic. GPRS reuses the existing GSM infrastructure to provide end-to-end packet-switched services. Benefits of GPRS include efficient radio usage, fast set-up/access time and high bandwidth with multiple time-slots. GPRS also provides a smooth path for GSM evolution to the third-generation mobile network. Specifically, a third-generation network can continue to utilise the GPRS–IP backbone network.

This chapter provides an overview of General Packet Radio Service (GPRS). The GPRS network architecture and its different entities are described here, and the interfaces among these entities with their major functionalities are provided. In addition to the GSM circuit-switched core network (CS-CN) for voice traffic, two new nodes, *SGSN: Serving GPRS Support Node* and *GGSN: Gateway GPRS Support Node,* are introduced. These two nodes constitute the packet-switched core network along with the other entities needed to handle packets. Mobility management and data transfer methods are described in connection with packet data services.

9.1 REVISITED GSM ADDRESSES AND IDENTIFIERS

Besides phone numbers and subscriber and equipment identifiers, several other identifiers have been defined in GSM. GSM distinguishes explicitly between user and equipment and deals with them separately. These identifiers are needed for the management of subscriber mobility and for addressing of all the remaining network elements.

IMEI International Mobile-station Equipment Identity It uniquely identifies a mobile station internationally allocated by the equipment manufacturer and registered by the network operator who stores it in the EIR (Equipment Identity Register). It is a kind of a serial number.

IMSI International Mobile Subscriber Identity It is used to uniquely identify the registered user. It is stored in the Subscriber Identity Module (SIM) (see Fig. 9.1). A mobile station can only be operated if a SIM with a valid IMSI is inserted into equipment with a valid IMEI.

The actual telephone number of a mobile station is the mobile subscriber ISDN number (MSISDN). It is assigned to the subscriber corresponding to one's SIM, such that a mobile station set can have several MSISDNs depending on the SIM.

TMSI Temporary Mobile Subscriber Identity It is assigned by VLR for the current location of a subscriber that has only local significance in the area handled by the VLR. It is stored on the network side only in the VLR and is not passed to the HLR.

9.2 GPRS NETWORKS ARCHITECTURE

The GSM components are changed as little to get economic packet data services. The standard GSM network has to be modified for transporting packet data in addition to normal voice communication over circuit-switch. As the general GSM architecture is already described in Chapter 8, only the new nodes and modification aspects of the GPRS network will be provided here. GPRS integrates a packet-based air interface into an existing circuit-switched GSM network. Figure 9.1 shows the general architecture for GPRS networks.

Some reservation and logical subdivision of certain GSM channels are required to integrate voice and data services in GPRS. Dynamic adaptation of channels is made in GPRS for load balancing among allocated channels within the respective cells.

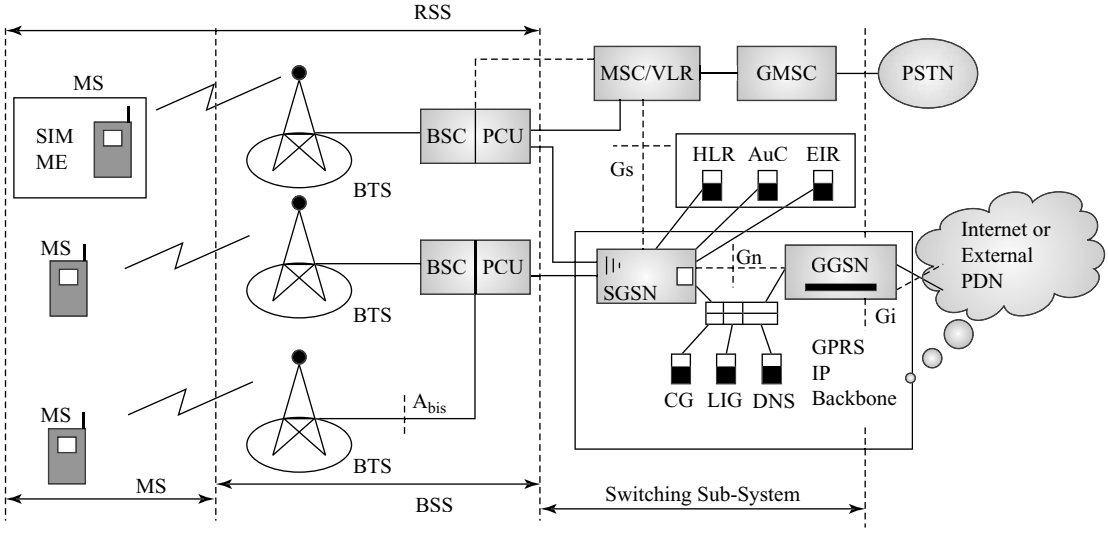


Figure 9.1 *GPRS network architecture*

9.2.1 Logical Architecture

For integration of voice and data in GSM networks, two new nodes are added that help in the routing of packet switched (PS) data in a separate parallel network. The first node is the node that serves the user managing mobility context and the second is required as the gateway to external data networks.

Serving GPRS Support Node (SGSN) The SGSN is basically the switching center-like MSC in GSM. SGSN is responsible for the delivery of data packets from and to the mobile stations within its service area. The location register of the SGSN stores the information of current cell and VLR along with the user profiles, i.e., IMSI addresses of all users for packet data network of GPRS under an SGSN.

The SGSN is the main entity for packet data network that handles mobility management, routing within the packet radio network, resource management, authentication and charging functions. Data packets addressed to SGSN are processed and mapped onto the IMSI value of the mobile. It also provides the authentication and encryption for the GPRS subscribers. SGSN establishes mobility management context for an attached mobile station (MS). Ciphering for packet-oriented traffic is also done by SGSN.

Gateway GPRS Support Node (GGSN) The GGSN is the interfacing node towards external PDNs (Packet Data Network) or the other PLMN (Public Land Mobile Networks). GGSN is capable of routing packets to the current location of the mobile. It has the access to the HLR in order to get current information of the mobile for terminating packet transfer. It converts the GPRS packets coming from the SGSN into the appropriate packet data protocol (PDP) format (e.g., IP or X.25) and sends them out on the corresponding PDN. In the other direction, PDP addresses of incoming data packets are converted to the GSM address of the destination user. The readdressed packets are sent to the responsible SGSN. For this purpose, the GGSN stores the current SGSN address of the user and his or her profile in its location register. The GGSN also performs authentication and charging functions. There are many relationships between the SGSNs and GGSNs.

SGSN connects GGSNs and other SGSNs through the IP network. Both of the GPRS Support Nodes (GSN) also collect charging data. The details of the volume of data are collected by the SGSN, whereas charging information detailing subscriber access to PDN is collected by the GGSN. The SGSN and GGSN functions may also be combined into one physical node. The communication between the GGSN and SGSN is done by a special protocol known as the GPRS Tunneling Protocol (GTP) that operates on the top of TCP/IP protocols.

It is to be noted that for extension of the existing GSM network by GPRS, the interfaces and reference points had to be redefined. The HLR, AuC and EIR may require minor modifications to support GPRS, which are generally in the form of software upgrade.

Charging Gateway (CG) CG generally takes the processing load off the SGSN and GGSN. It also provides single logical links for the billing system.

Lawful Interception Gateway (LIG) This is required in many countries for monitoring traffic for the law-enforcement agencies that require court order. When data packets traverse the GPRS network, it may be intercepted and forwarded to the agencies.

Domain Name System (DNS) When GPRS mobile users wish to get connection to any external network through GGSN, it first selects an APN (Access Point Name) from a list in the mobile device. DNS is required to get the correct GGSN according to the IP address of the GGSN. The SGSN resolves the APN to the correct GGSN IP address. The APN may be the name in the form of text, such as Internet, Intranet or home network. For any network, it is common practice to use 'net;' and for any wireless access point, it is 'WAP'. WAP is a connection to the wireless access point, gateway. Different operators, who will charge differently for different access points and Internet access, may operate external networks.

Border Gateway (BG) The border gateway is used as the gateway to a backbone network connecting different operators together, and an IP router. It is generally implemented as the same hardware platform as GGSN. This backbone basically supports inter-PLMN roaming facility with the help of Global Roaming Exchange (GRX). The operation and configuration of this connection follows the roaming agreements across the network operators.

Packet Control Unit (PCU) The GPRS cell phone will transmit data in Packet Switched (PS) mode and the voice will be transmitted in Circuit Switched (CS) mode. So there needs to be a network to differentiate the different kinds of calls and send them to the respective core network—voice calls to the MSC and data calls to the SGSN. The network unit called Packet Control Unit (PCU) does this. Generally, the PCU is placed at the BSC site, though it may be placed at the site of the BTS or before the switch.

The PCU performs the following functions:

1. Packet segmentation and reassembly on the uplink and downlink
2. Access control
3. Scheduling for all active transmissions including radio channel management
4. Controlling transmission for checking, buffering and retransmission

9.2.2 Classes of GPRS Equipments

As GPRS users need to handle both voice and data services, the following classifications are made in the perspective of GPRS equipments.

Class A Equipments that handle voice calls and transfer data at the same time.

Class B Equipments that can handle voice or data traffic separately, and can put a packet transfer on hold to receive a phone call.

Class C Equipment that can handle both voice and data, but has to be disconnected from one mode explicitly in order to enable the other.

9.2.3 GPRS Interfaces and Reference Points

For extension of a GSM network, certain interfaces and reference points had to be redefined. Figure 9.2 shows the interfaces between the new network nodes and the GSM network as defined by ETSI.

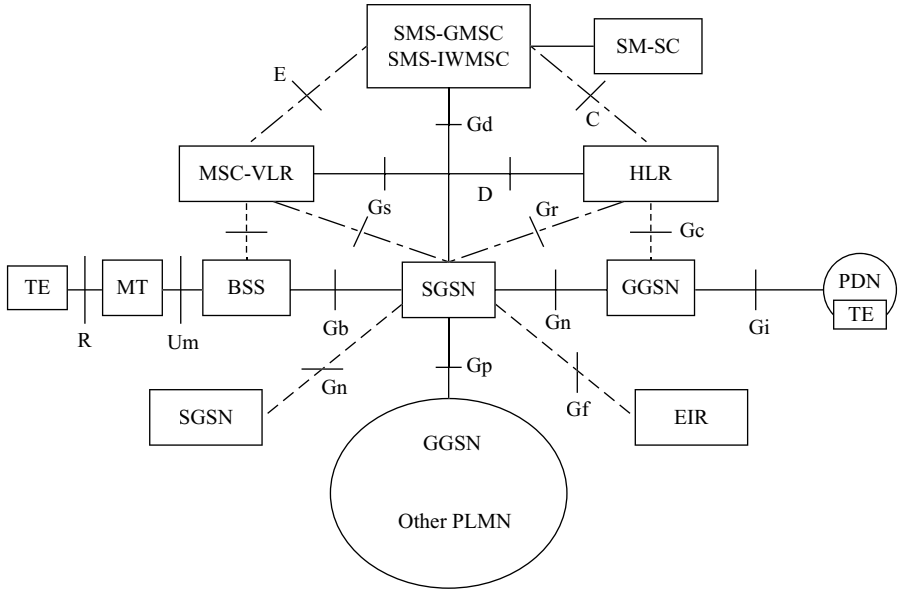


Fig. 9.2 *GPRS network interfaces*

Interfaces

Gb between the BSS and SGSN is to transport both signaling and data traffic, and is based on frame relay protocol.

Gc works between HLR and GGSN and provides the GGSN with the access to subscriber information. The protocol used here is MAP, and the interface is used for signaling only.

Gd interface connects the SGSN to an SMS gateway, thus enabling the SGSN to support SMS services.

Gf interface connects the SGSN to EIR and allows the SGSN to check the status of a particular mobile device, that is whether it is stolen or approved for connection. Across the Gf interface, the SGSN may query the IMEI of a mobile station trying to register with the network.

Gi is the reference point rather than an interface, and refers to the connection between the GGSN and some external network: IPv4, IPv6 and PPP are supported by GPRS and the Gi interface.

Gn interface resides between the GPRS support nodes. It consists of a protocol stack that includes IP and GTP. GTP has two parts, the GTP-U, which is used to carry user data, and the GTP-C that is used to carry control data. The Gn interface will be used if SGSN and GGSN are located in the same PLMN, (public land mobile network), whereas the Gp interface will be used if they are in different PLMNs. The Gn and Gp interfaces are also defined between two SGSNs. This allows the SGSNs to exchange user profiles when a mobile station moves from one SGSN area to another.

All GSNs are connected via an IP-based GPRS backbone network. Within this backbone, the GSNs encapsulate the PDN packets and tunnel them using the GPRS Tunneling Protocol GTP. There are two kinds of GPRS backbones:

The intra-PLMN backbone networks connect GSNs of the same PLMN and are therefore private IP-based networks of the GPRS network provider.

The inter-PLMN backbone networks connect GSNs of different PLMNs. A roaming agreement between two GPRS network providers is necessary to install such a backbone.

Gr works between the SGSN and HLR, providing the SGSN with access to subscriber information. The protocol used is MAP (Mobile Application Part), and it is used for signaling. The HLR stores the user profile, the current SGSN address, and the PDP addresses for each GPRS user in the PLMN. When the MS registers with a new SGSN, the HLR will send the user profile to the new SGSN. The signaling path between GGSN and HLR via the Gc interface may be used by the GGSN to query a user's location and profile in order to update its location register.

Gs is the optional interface used for signaling between SGSN and VLR that is collocated with MSC. It uses BSSAP + protocol, a subset of BSSAP (Base System Application Part Protocol) to support the signaling between the SGSN and MSC/VLR. This interface enables efficiency by coordinating signaling to the mobile device with combined LA (Location Area) and RA (Routing Area) update and IMSI attach and detach procedures. Apart from the G interfaces, two other U interfaces are there across the air for both GPRS and UMTS networks:

Um interface is the modified GSM air interface between the mobile device and the RAN for GPRS networks and Uu is the UMTS air interface between the mobile device and the fixed network, which provides GPRS services.

9.2.4 GPRS Logical Channel

Table 9.1 shows the logical channels in the GPRS network, which are divided into two categories: traffic channels, and signaling or control channels. Logical channels are defined to perform multiple functions like broadcast, paging, signaling, synchronisation and payload transport.

Table 9.1 *Logical channel in GPRS networks*

Group	Channel	Function	Direction
Packet Broadcast control channel	PBCCH	Broadcast control	BSS → MS
Packet data traffic channel	PDTCH	Data traffic flow	MS ↔ BSS
Packet dedicated control channel	PACCH	Associated control	MS ↔ BSS
	PTCCH	Timing advance control	MS ↔ BSS
Packet common control channel	PRACH	Random access	MS ↔ BSS
	PAGCH	Access grant	MS ↔ BSS
	PPCH	Paging	MS ↔ BSS
	PNCH	Notification	MS ↔ BSS

PDTCH The packet data traffic channel is employed for the transfer of user data, assigned to one mobile station (MS) during data transfer. Each MS can use several PDTCHs simultaneously. The packet broadcast control channel (PBCCH) is a unidirectional point-to-multipoint signaling channel from the base station subsystem (BSS) to the mobile stations. The packet timing advance control channel (PTCCH) is used for adaptive frame synchronization. In GPRS, the coordination between circuit-switched and packet-switched logical channels is very important. The packet common control channel (PCCCH) is a bi-directional point-to-multipoint signaling channel that transports signaling information for network access management, e.g., for allocation of radio resources and paging. It consists of four subchannels as given in Table 9.1.

If the PCCCH is not available in a cell, a mobile station can use the common control channel (CCCH) of conventional GSM to initiate the packet transfer. Moreover, if the PBCCH is not available, it will listen to the broadcast control channel (BCCH) to get informed about the radio network.

The packet associated control channel (PACCH) is always allocated in combination with one or more PDTCH that are assigned to one mobile station. It transports signaling information related to one specific mobile station, i.e., power control information.

9.2.5 GPRS Service Types

A PLMN provider is responsible for the data transfer between the GPRS Service Access Point (SAP) in the fixed network and the GPRS SAP in the MS. Two kinds of services are present: (i) Point-to-Point (PTP), and (ii) Point-to-Multipoint (PTM) services.

PTP is used for single packet transfer between two subscribers. It is operated both in connection-oriented mode as well as in connectionless mode. The first one is the Connection Oriented Network Service (CONS) for X.25, while the second one is the Connectionless Network Service (CLNS), i.e., for IP.

PTM supports the transmission of data packets between a service user and a specified group inside a certain geographical region. It is again divided into PTM-Multicast (PTM-M). Data packets are broadcast over a certain geographical region. A group identifier indicates whether the packets are intended for all users or to a group of users. IP Multicast (IP-M) is used to realise this service.

PTM-G is meant for only a group of users (PTM-Group). The messages are addressed explicitly to a specified group and are sent in all geographical regions where the group members are located.

9.2.6 Parallel Use of Services

During the GPRS session, circuit-switched connection may also be possible both for voice and data. Again it is also possible to send and receive GPRS data while carrying out a telephone call. PTP and PTM services are provided for the parallel services.

Apart from these services, some more additional services like SMS are possible both for Mobile Originated Calls (MOCs) or Mobile Terminated Calls (MTCs). Depending on the current load situation, the transfer of an SMS may be delayed.

9.3 GPRS SIGNALING

The GPRS protocol architecture is mainly divided into two planes—user plane (also known as *transmission plane*) and control plane (also known as *signaling plane*). When packet data transmission is requested both in uplink or downlink, the user plane handles this transmission. This would require the user plane to have knowledge of the addresses of the peer entities, and the status of network elements prior to the session initiation, and it has to be kept up-to-date during an *ongoing* session.

On the other hand, GPRS control plane handles the network management functionalities like GPRS Mobility Management (GMM), Session Management (SM) and Quality of Service (QoS).

The GMM tracks the MS's location during the movement, updates its databases with new location information and supports cell-change procedures.

The SM manages the logical context between the MS and the external PDNs when a new session is set up or for changing its performance requirements. The QoS management in GPRS networks supports the network to establish QoS issues that are based on the SM procedures.

In the following section, we will discuss about session management, mobility management, and routing of GPRS network. It will help to understand how an MS registers first with the GPRS network and identifies it by the external PDNs, how packets are routed to or from MSs, and how networks keep track of current location of the MS.

9.3.1 GPRS Mobility Management Procedures

The GMM comprises access control and authentication functions. To obtain the access to the GPRS network, the MS has to perform a GPRS attach procedure. During Mobility Management (MM) procedures, user data can be transmitted while signaling is going on. This may lead to the loss of data during attach, authentication and location update processes.

GPRS Attachment and Detachment Procedures Before a mobile station can use GPRS services, it must register with an SGSN of the GPRS network. The network checks if the user is authorized, copies the user profile from the HLR to the SGSN, and assigns a Packet Temporary Mobile Subscriber Identity (P-TMSI) to the user. This procedure is called GPRS attach. MS communicates with the SGSN during the attach procedure. The disconnection from the GPRS network is called GPRS detach. It can be initiated by the mobile station or by the network (SGSN or HLR). Figure 9.3 shows the steps for attach procedure.

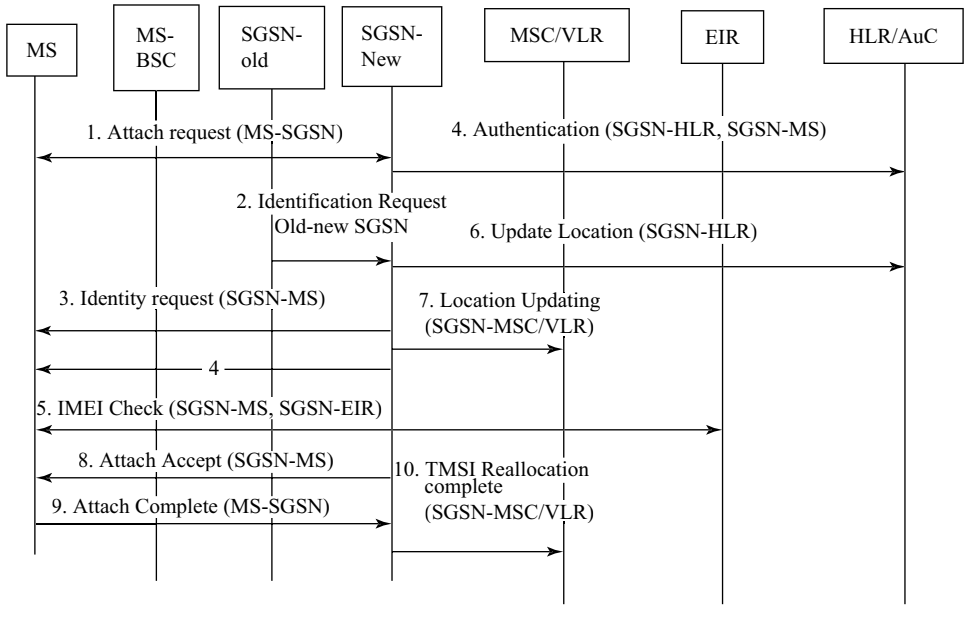


Fig. 9.3 *Overview of GPRS attach procedure*

The messages concerning the GPRS attach procedure between MS and the SGSN belong to the protocol GPRS mobility management GMM.

The parameter of the Attach Request message includes the MS's GPRS multi-slot capabilities, GPRS ciphering algorithms, the type of attach performed and further modes and capabilities of the MS. If the MS is identified by P-TMSI (Packet-TMSI) and SGSN is changed, the GTP-C (used for signaling tunneling) of the new SGSN sends the identification request message to the old SGSN for the MS IMSI. In reply, the old SGSN sends IMSI and authentication parameters. In case of the unknown MS, both of old and new SGSNs, the GMM entity of the SGSN sends the identity request to MS which in reply provides the IMSI. Security functions are performed after authentication between the SGSN and GPRS registers.

Location update is done through the update message using MAP (Mobile Application Part). Acknowledgement between SGSN and HLR or VLR is done using the BSSAP+ signaling protocol (discussed later). After that, the GMM entity of the SGSN sends the Attach Accept message with the assigned P-TMSI and the VLR-TMSI to the MS. If those values are changed then the MS acknowledges the received TMSI by sending the Attach Complete message to the SGSN. On the other hand, if the VLR-TMSI is changed then the BSSAP+ entity of the SGSN confirms the VLR-TMSI reallocation by sending a TMSI reallocation complete message to the VLR. If the attach Request cannot be accepted, the GMM entity of the SGSN returns the Attach Reject message to the MS.

GPRS Detach Procedure Through the GPRS detach procedure, the MS informs the network that it does not want to access the GPRS services, and also the network can inform the MS that it does not have access to the SGSN-based services any more. Two possible detach modes are *explicit* (either network or MS explicitly requests the detach procedure) and *implicit* (the network detaches the MS without notification).

9.3.2 Session Management and PDP Context

After a successful GPRS attach, the MS must apply for one or more addresses used in the external PDNs to exchange data packets, e.g., for an IP address in case the PDN is an IP network. This address is called PDP address (Packet Data Protocol address).

Two entities in the core network are necessary to provide the mobility and session management functionalities such as routing area update and PDP context activation. These are HLR and DNS (Domain Name Server). As HLR is a part of the NSS, the DNS is a new entity to be implemented. As DNS is purely for the packet-switched core network, the HLR is a unit that offers services both for MSC and SGSN. It is basically a workstation which is directly connected to the core network and addressed via an IP address (refer to Fig. 9.4). This address can be found out in the configuration of every SGSN.

An important task of DNS is the address resolution for GGSN IP address during the process of PDP context activation. The MS sends a request to the SGSN with the domain name of the network it wants to be connected. The network is known as Access Point Name (APN), a logical name like an Internet address (www.jdvu.edu.in) with a difference that APN does not address the application server. Each APN is connected through the GGSN to a particular SGSN to which the mobile sends request. For each active PDP context, the SGSN finds the IP address of GGSN to establish the path (GTP tunnel) between SGSN and GGSN for packet transfer. So, in reply to the request with a specified APN, an IP address of the GGSN is sent back to the SGSN that connects to the specified network with that APN.

For each session, a PDP context is created which describes the characteristics of the session. It contains the PDP type (e.g., IPv4, IPv6, X.25 or PPP), the PDP address assigned to the mobile station or an APN (i.e., intranet.edu.in), the requested QoS, and the address of a GGSN that serves as the access point to the PDN. This context is stored in the MS, the SGSN, and the GGSN. As a result of an active PDP context, the whole GPRS network knows how to route the IP packets of the GPRS service. The external PDN is able to send and receive the data packets. The path from the MS to the GGSN via the Gi interface is well defined after the activation of the PDP context.

The message concerning the activation of the PDP context between the MS and SGSN belongs to the protocol Session Management—SM. A transaction identifier TI is allocated to an activated PDP context. The TI identifies the session on the SM layer as shown in Fig. 9.5.

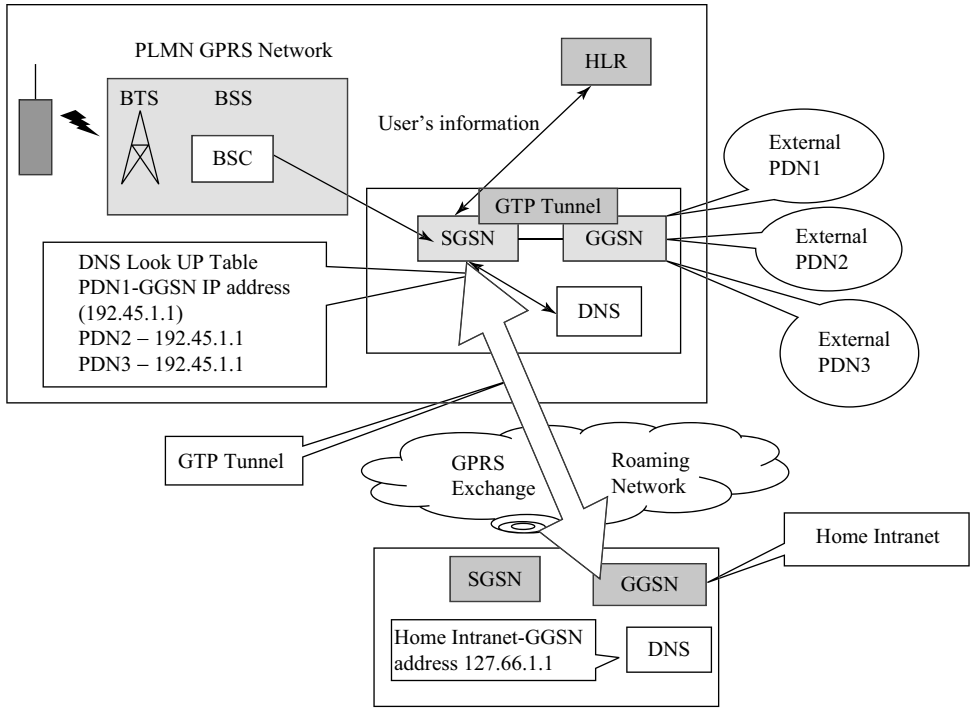


Fig. 9.4 *Illustration for PDP context*

The address assignment for PDP context may be static or dynamic. For a static case, the home PLMN (Public Land Mobile Network) of the MS permanently assigns the PDP address whereas in a dynamic case, the address assignment is done upon activation of PDP context by the operator's of the user, either in home PLMN or in visited PLMN. GGSN is responsible for the allocation and the activation/ deactivation of the PDP address allocation for a dynamic case. This address is used for packet forwarding between the Internet and the GGSN, and within the GGSN it is used to tunnel user data through the core network (SGSN to GGSN).

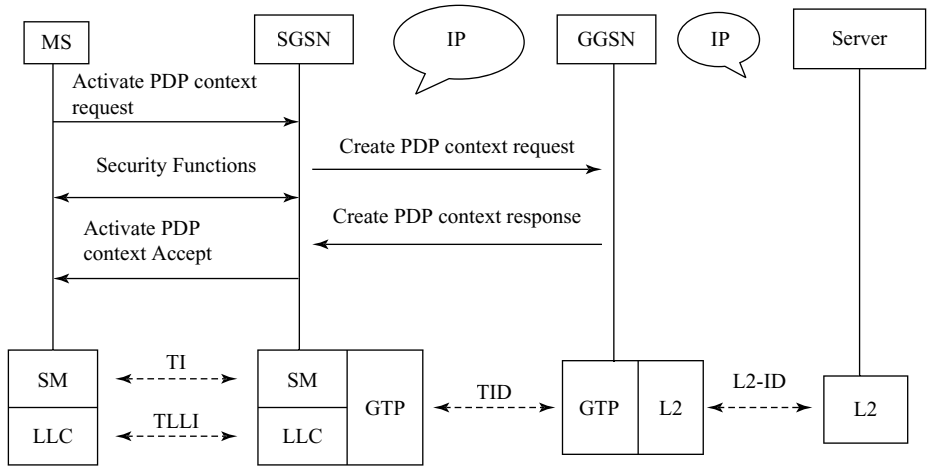


Fig. 9.5 *Steps for PDP context activation procedures*

Figure 9.5 shows the active PDP context procedure. With the message activate PDP context request, the MS informs the SGSN for the PDP address within the fields (PDP type, PDP address, QoS requested, access point). For dynamic assignment, the address field will be empty. After that, security functions are performed to authenticate the user requesting the service. For a valid user if the PDP context request is granted, the SGSN sends the create PDP context request to the GGSN. The GGSN creates a new entry to the PDP context table that enables the GGSN to route data packets between the SGSN and the external PDN. After that, the GGSN sends the PDP context response to the SGSN that contains a new PDP address. The SGSN updates the PDP context table and sends the new PDP context accept message to the MS.

The information for the activated PDP context is sent over the Gn interface with the GTP which uniquely identifies the PDP context between the SGSN and GGSN with a Tunnel Identifier TID. Each IP packet transmitted from the MS is put into an envelope that carries two identifiers. One is TFI—Traffic Flow Identity and the second is NSAPI—Network Service Access Point Identifier. TFI identifies the active GPRS MS in a certain cell, and the NSAPI identifies one of the activated PDP contexts of GPRS MS.

The Logical Link Protocol (LLC) enables information to be transferred between the MS and SGSN. The LLC also ciphers the connection between the MS and the SGSN. This logical connection is defined by the Data Link Connection Identifier (DLCI), which contains the Temporary Logical Link Identifier (TLLI) and the Service Access Point Identifier (SAPI). After receiving packets, the PCU translates TFI into TLLI and forwards the packets to the SGSN. The TLLI remains same unless the SGSN allocates the new one since the GPRS attach procedure, i.e., the TLLI is a unique identifier of the serving MS in the SGSN. The TLLI is generally derived from P-TMSI (Packet Temporary Mobile Subscriber Identifier). The TLLI is used as an identifier in LLC protocol whereas the P-TMSI is used in GMM (GPRS Mobility Management) protocol.

It is to be mentioned that the GGSN sets the correct physical port at Layer 2 on the Gi interface to this PDP context to route the IP packets to the external PDN. Layer 2 may be Ethernet, Frame relay or ATM.

For IPv4, the MS has no IP address, and so a temporary IP address is provided to the MS from the external IP network and GGSN is responsible to do this, i.e., it sends a request for the IP address to the IP network.

9.3.3 Data Transfer Through GPRS Network and Routing

After the completion of PDP context activation, a logical connection between the MS and the external IP network is made. The MS gets a temporary IP address from the external PDN and is seen as being a member of the IP network.

The MS is identified by the TLLI on the LLC layer by the SGSN and the PDP context is identified by NSAPI (Network Service Access Point Identifier) on the SNDCP (Sub-Network Dependent Converged Protocol) layer. Between the SGSN and GGSN, the MS and its PDP is identified by TID (Tunnel Identifier).

On the Gi interface, the GGSN sets the right physical port on Layer 2 to this PDP context. Thus, the routing of incoming and outgoing IP packets for a particular PDP context is perfectly defined.

The data packets transported between MS and PDN are compressed and segmented between the MS and the SGSN. This task is done by the SNDCP. In the downlink direction (PDN to MS), the SNDCP compresses and segments the IP packets between the MS and SGSN transmissions. In the uplink, it reassembles and decompresses the LLC packets to get the IP packet again.

Let us consider that the packet data network is an IP network. A GPRS MS located in PLMN1 sends IP packets to a host connected to the IP network, e.g.,to a Web server connected to the Internet. The concerned SGSN with which the MS is currently attached, encapsulates the IP packets coming from the MS, examines the PDP context, and routes them through the intra-PLMN GPRS backbone to the appropriate GGSN. The GGSN decapsulates the packets and sends them out on the IP network, where IP routing mechanisms are used to transfer the packets to the access router of the destination network. The latter delivers the IP packets to the host.

Figure 9.7 shows two intra-PLMN backbone networks of different PLMNs connected with an inter-PLMN backbone. It also helps to explain the routing procedure. The gateways between the PLMNs and the external

inter-PLMN backbone are called Border Gateways (BGs). Among other things, they perform security functions to protect the private intra-PLMN backbones against unauthorized users and attacks.

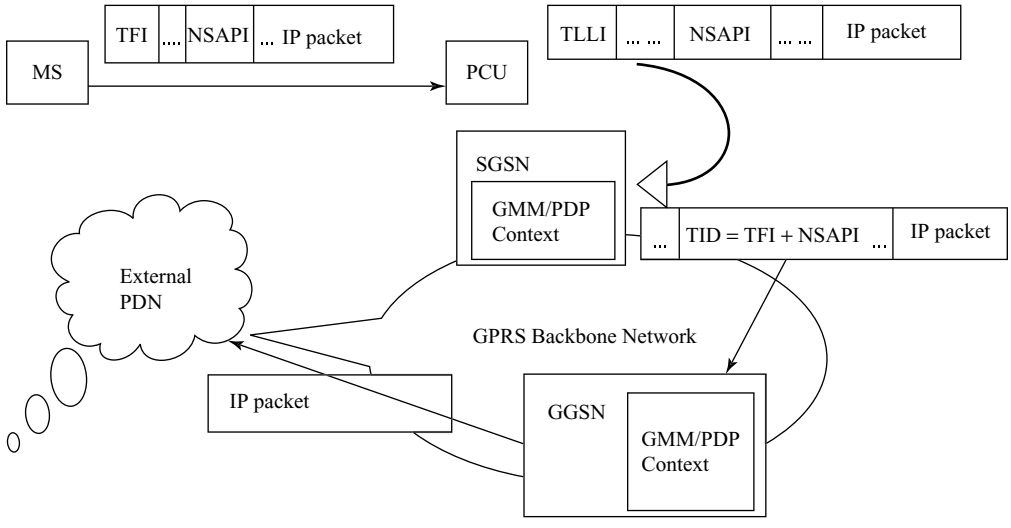


Fig. 9.6 *Transfer of data through GPRS network*

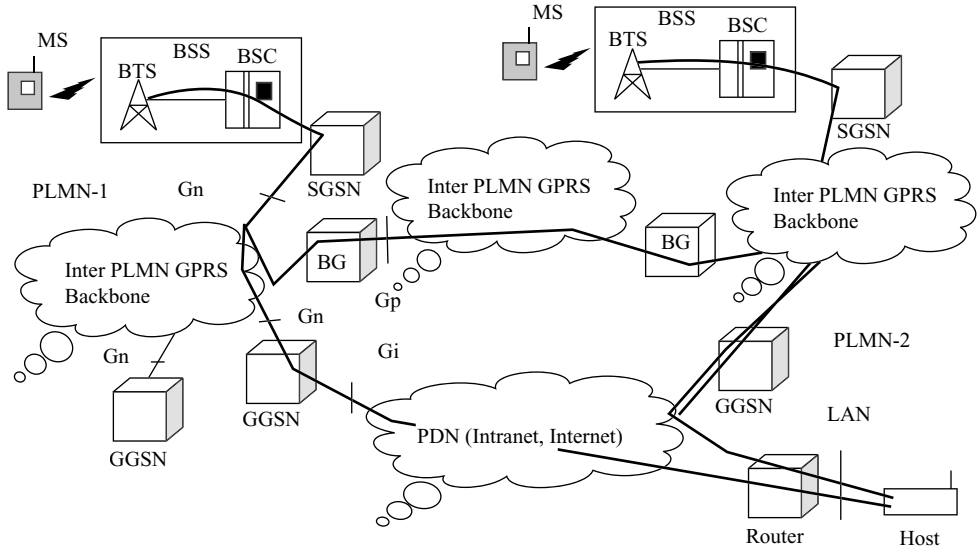


Fig. 9.7 *Example GPRS network for routing*

Let us now assume that the home-PLMN of the MS is PLMN2. The GGSN of PLMN2 has assigned an IP address to the mobile. Thus, the MS's IP address has the same network prefix as the IP address of the GGSN in PLMN2. The correspondent host is now sending IP packets to the MS. The packets are sent out onto the IP network and are routed to the GGSN of PLMN2 (the home-GGSN of the MS). The latter queries the HLR and obtains the information that the MS is currently located in PLMN1. It encapsulates the incoming IP packets and tunnels them through the inter-PLMN GPRS backbone to the appropriate SGSN in PLMN1. The SGSN decapsulates the packets and delivers them to the MS [2].

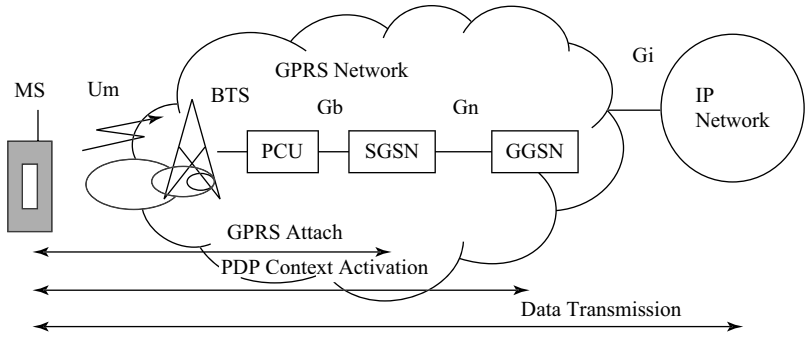


Fig 9.8 *GPRS procedures*

Figure 9.8 is the illustration of the combined process of GPRS procedures—attach, PDP Context and Data transfer.

9.4 GPRS STATES OF MOBILITY MANAGEMENT

GPRS Mobility Management (GMM) is based on three different Mobility Management (MM) states defined at the MS and at the SGSN. These are

1. Idle state
2. Standby state
3. Ready state

In each of the MM states, the MS and the SGSN hold the different information about the GPRS terminal—known as MM context.

In the **idle state,** the MM context is empty, as the subscriber is not attached to the GPRS network. No valid information for the MS in which cell and routing area (RA) resides is present. Therefore, no MM procedure is performed. The MS is seen to be not reachable to the network.

In the **standby state,** the MS is attached to the network. Each of the MS and SGSN have established an MM context. The MS and SGSN know RA of the MS. A paging message is sent by SGSN to all cells within the routing area.

In the **ready state,** the MS position is known to SGSN up to the cell level. The SGSN can therefore send continuous data packets on the downlink even if it changes the position of the cell supported by Location Management (LM).

In the idle state, no location updating is performed, i.e., the current location of the MS is unknown to the network. An MS in the ready state informs its SGSN of every

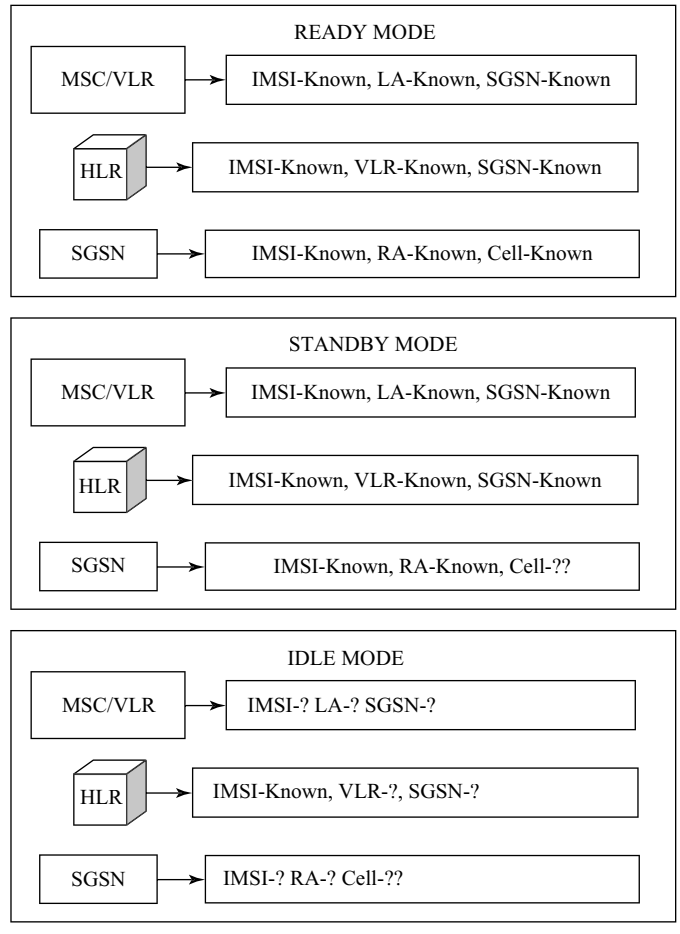


Fig. 9.9 *Different state conditions for MM*

movement of a new cell. For the location management of an MS in the standby state, a GSM location area (LA) is divided into several routing areas (RA). In general, an RA consists of several cells. The SGSN will only be informed when an MS moves to a new RA; and the cell changes will not be disclosed. To find out the current cell of an MS in the standby state, the paging of the MS within a certain RA must be performed. For MSs in the ready state, no paging is necessary. Figure 9.9 shows the different state conditions of the MM.

9.4.1 GMM State Transitions

All possible state transitions for MM is shown in Fig. 9.10. When the MS transits from the idle state to the ready state, it performs GPRS attach procedure and a logical link is established with the SGSN. When the implicit detach procedure is set up, the MM and PDP context from the SGSN goes to the idle state, and the GGSN PDP context is deleted. When the MS sends a PDU to the SGSN, the MM state in SGSN switches to ready and after receiving the PDU by the SGSN, it returns to the ready state. The SGSN can also force the standby state before the ready state timer expires.

There are three types of timers in the GMM that controls the state transitions:

1. The Ready timer controls the duration of an MM context that remains in the ready state in the MS and in the SGSN. This timer is reset and begins running on the MS side when the LLC PDU is transmitted on the MS side, and the LLC PDU is correctly received on the SGSN side.
2. The periodic RA (Routing Area) update timer is set periodically and its length, which is constant in one RA, is included in the RA Update Message from the SGSN. At expiry, the MS performs an RA update procedure.
3. The mobile reachable timer controls the periodic RA update procedure in the SGSN, which is slightly longer than the RA update timer. On expiry, the MS is out of network coverage.

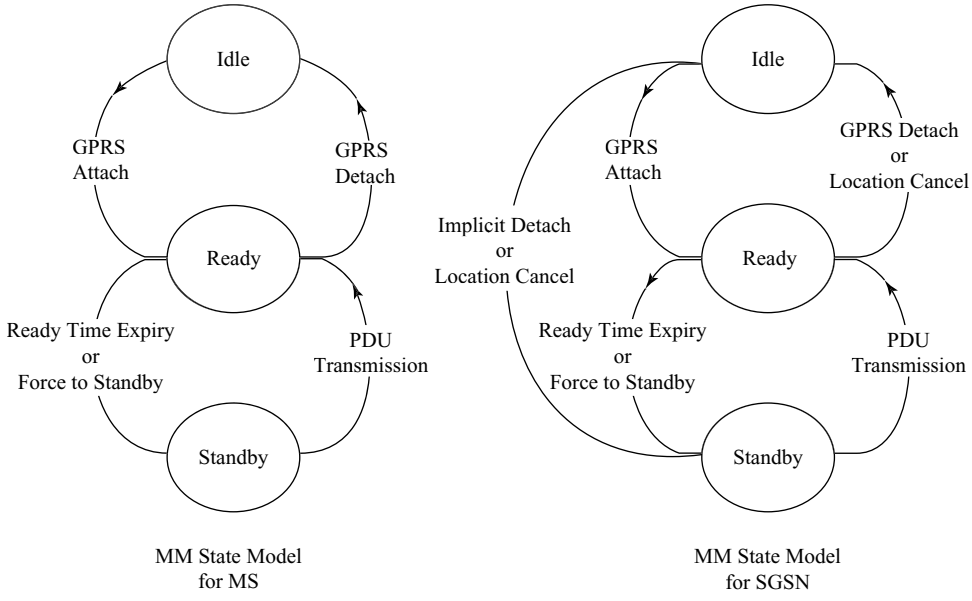


Fig. 9.10 *GPRS MM state transition*

9.5 GPRS LOCATION MANAGEMENT PROCEDURES

The main task of location management is to keep track of the user's current location, so that incoming packets can be routed to his or her MS. For this purpose, the MS frequently sends location update messages

to its current SGSN. If an update were performed every time the MS changed cells, a very large amount of signaling would result. On the other hand, if the position of the MS is not known, a paging message is to be sent to the entire mobile network during the incoming call arrival, again resulting in large signaling overhead requirement. So a method to minimise the signaling is required.

In circuit-switched GSM network, the group of cells where the location identification (ID) for the MS remains the same is known as Location Area (LA), whereas in GPRS networks it is called the Routing Area (RA). When a subscriber receives a call, it is paged in all the cells belonging to the group. The location area ID (LAI) for each mobile is stored in the VLR, and the Routing Area ID (RAI) for each mobile is stored in the SGSN Location Register (SLR). When a user moves from cell to cell under the same LA or RA, no location update is required. If the user moves under different LA or RA, the VLR is informed about the change. If the user moves to a new VLR area, the HLR is also informed about this change and the record in the previous VLR is erased.

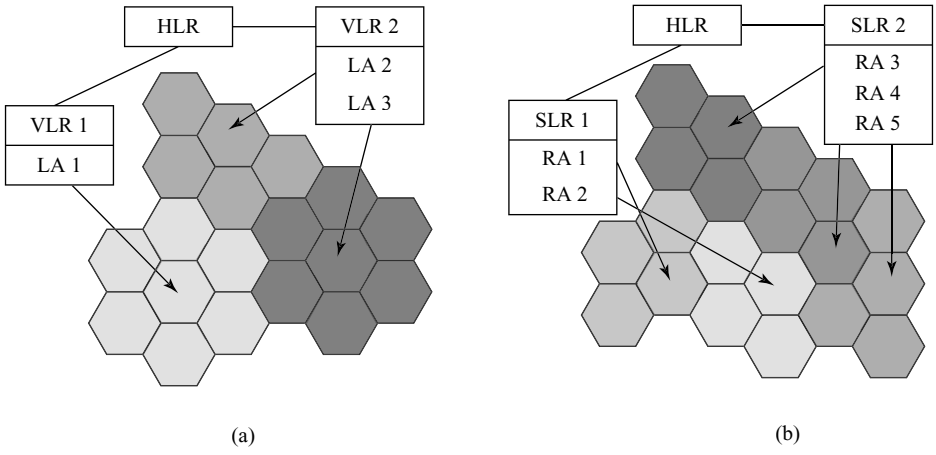


Fig. 9.11 *(a) Location Area for GSM, (b) Routing Area in GPRS Networks*

The hierarchical LA /RA structure is shown in Fig. 9.11. In case of packet-switching network, the total amount of signaling may be large due to the large number of paging requirement during downloading for the roaming user. As a consequence, the number of cells in a group for the GPRS network is small. Thus, RA size is smaller than LA size. For GMM, it is important that every RA entirely lies with one LA, i.e., one LA may contain one or more routing area.

The Location Management Procedure is described when the MS moves from one cell to another cell and remains only in the standby or ready mode. Three following scenarios are possible:

1. MS initiates cell update procedure.
2. MS initiates Routing Area Update.
3. MS initiates combined Location area and Routing Area Update.

9.5.1 Cell Update

It is required when the MS moves from one cell to another under the same MSC or SGSN as shown in Fig. 9.12. The MS performs cell update procedure by

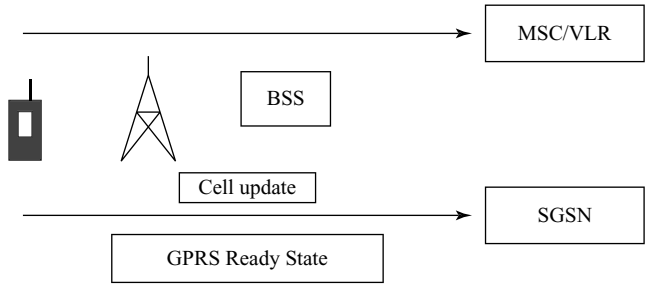


Fig. 9.12 *GPRS cell update process*

sending a message containing the P-TMSI to the SGSN. The PCU adds the Cell Global Identity (CGI) while moving towards SGSN and records this change of cell. Traffic is directed to this new cell.

9.5.2 Routing Area Update

If the MS moves into new RA then RA update is required. The mobile device detects the new RA identifier RAI, or the RA timer may expire in case of periodic update. The RA update may be Intra SGSN or Inter SGSN.

Intra SGSN RA Update The MS moves to a new routing area controlled by the same SGSN. The new RA is controlled by the same SGSN where the MS is moved. The SGSN knows the MS location and there is no need to inform HLR or the GGSN. The MS is given a new P-TMSI. Figure 9.13 shows the Intra-SGSN routing update.

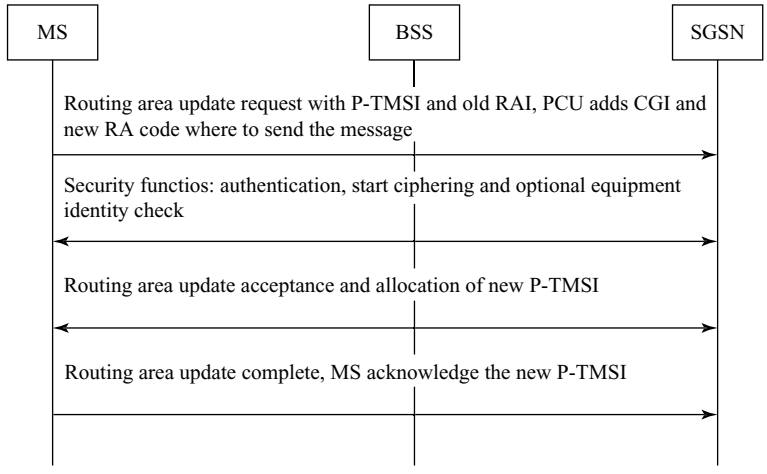


Fig. 9.13 *Intra-SGSN routing area update*

Inter-SGSN RA Update The MS is moved to a new RA under a different SGSN. The inter-SGSN RA update procedures involve the GGSNs and GPRS registers, HLR and VLR. The new SGSN has to store the subscriber information and RAI. The old SGSN forwards the data packets buffered during the changeover period. The following steps are performed for inter SGSN RA update and are illustrated in Fig. 9.14.

1. After reaching the new SGSN, the MS sends the RA update request to the new SGSN with the information of the old RAI and the old P-TMSI. The BSS adds the identity of the new cell before passing the message to the SGSN.
2. Authentication procedure is done at the new SGSN, and it sends the SGSN context request to the old SGSN to get information about MM and PDP contexts for the MS.
3. For a valid user, the old SGSN responds to the new SGSN with SGSN context response message. It stops sending packets directly to the mobile and starts a timer.
4. Security functions concerning the ciphering are executed.
5. The new SGSN sends an acknowledgement to the old SGSN to notify its readiness to accept packets destined for the mobile device from any active PDP context.
6. The old SGSN starts sending packets to the MS via the new SGSN. Packets are tunneled from the GGSN to old SGSN and then to the new SGSN.

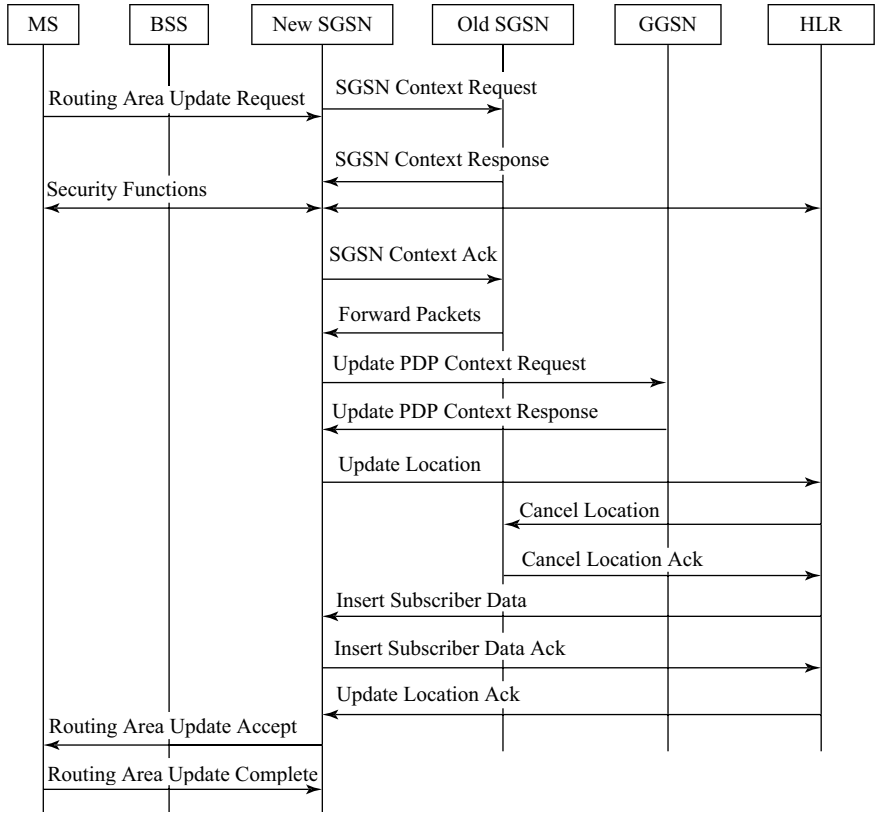


Fig. 9.14 *Inter-SGSN routing area update process*

7. The new SGSN informs the GGSNs about the new position of the mobile and sends requests for updating the PDP context with the information of the new SGSN address, TEID (Tunnel End Identifier) and the negotiated QoS.
8. The GGSN updates the database and replies with an update PDP context response.
9. The new SGSN informs the HLR of the SGSN change by sending an MAP update location message to the HLR that includes the SGSN address and the IMSI value of the MS.
10. The HLR MRP entity sends a cancel location message to the old SGSN. The old SGSN removes the MM and PDP contexts after a timeout.
11. The old SGSN acknowledges the cancel location message.
12. The HLR sends an Insert Subscriber Data message to the new SGSN. This includes the IMSI and the GPRS subscription data.
13. In reply, the new SGSN sends the acknowledgement for the Insert Subscriber Data.
14. The HLR acknowledges the location update by sending update location acknowledgement to the new SGSN.
15. The new SGSN now establishes the logical link connection with the MS and sends the routing area update accept message.
16. The MS acknowledges the new P-TMSI with an RA update complete message. It confirms the reception of LLC PDUs those were additionally forwarded from the old SGSN, and then these LLC PDUs are discarded by the new SGSN.

9.6 GPRS ROAMING

Users under GPRS networks want the roaming services to use their same terminal worldwide in a secure, reliable and economic way. So, GPRS operators need to provide this service among networks satisfying those conditions. The perspective of roaming is to get Internet services at the roaming country, accessing email services and also the intranet service of corporate users. Some prerequisites are required to facilitate roaming as given under:

1. The operator for the home PLMN must have roaming agreement with the operator in the country the user is visiting.
2. The mobile phone of the user must be GSM compatible in the visited location and at the home PLMN.
3. The user must be entitled for roaming services that is stored in the HLR database.

GPRS roaming is more difficult than GSM because of the billing complexity that is measured on the volume of data transferred. However, GPRS roaming exchange (GRX) service is the solution used by operators. Suppose a roaming user wants to connect to the local SGSN in the visited country. Before getting this access, the visited PLMN must contact the home PLMN to authenticate the user and subscribed services. The SGSN then only provides Internet access via the local GGSN.

Consider a typical case when the user is geographically remote from the network service provider. The user's home network under a GGSN is far away from the SGSN where the mobile user is presently situated. The operators may take the advantage of a public network such as the Internet or VPN (virtual private network) services. The VPN may be able to share common underlying infrastructures with other IP services that would provide a secure IP network dedicated to the GPRS roaming services.

The basic entities of a GRX network are DNS, Border Gateway and IP based transport between the different GPRS networks, as shown in Fig. 9.15. The Customer Edge Router (CER) performs the boarder gateway functionality for the GPRS operator's network, and the Provider Edge Router (PER) provides the boarder gateway functionality of the GRX network. In this way, several GRX networks can be interconnected to support wider mobility of the users.

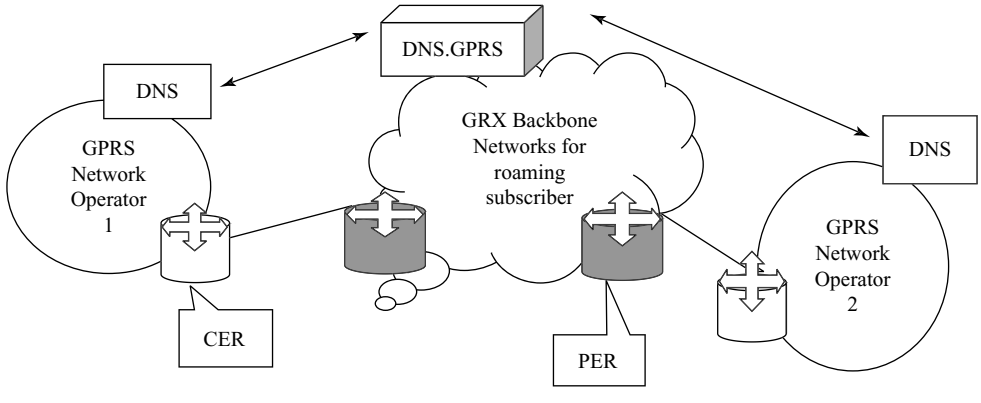


Fig. 9.15 *GPRS roaming exchange network*

The advantage of such a network is a unique agreement throughout the GRX and all connected GPRS PLMNs, reducing complexity, increasing ease of maintenance and enhancing the billing model. The GRX can take care of billing record and charging providing transparent service to all concerned third parties and operators.

Data transfer through GRX networks can be explained with Fig. 9.16. Suppose a roaming user is in the visited network, and is registered with the SGSN. He wants to access an intranet of a corporate office

that can only be accessed by his home network. In this situation, a connection between the home and visited networks of the user is to be established via GRX. The MS first sends the request for GPRS access with its APN to the visited SGSN. The SGSN interrogates the local DNS with the requested APN, and fails to recognize it. This request is passed to the higher level DNS or to the GRX DNS. If it is recognized by GRX DNS, it will return to the visited DNS, otherwise it will be forwarded to the home DNS. The visited SGSN obtains an IP address of the required GGSN, and requests PDP context activation to the GGSN. The GGSN accepts PDP contexts and a GTP tunnel is established through which a two-way data communication is made between the MS and the destination via GGSN.

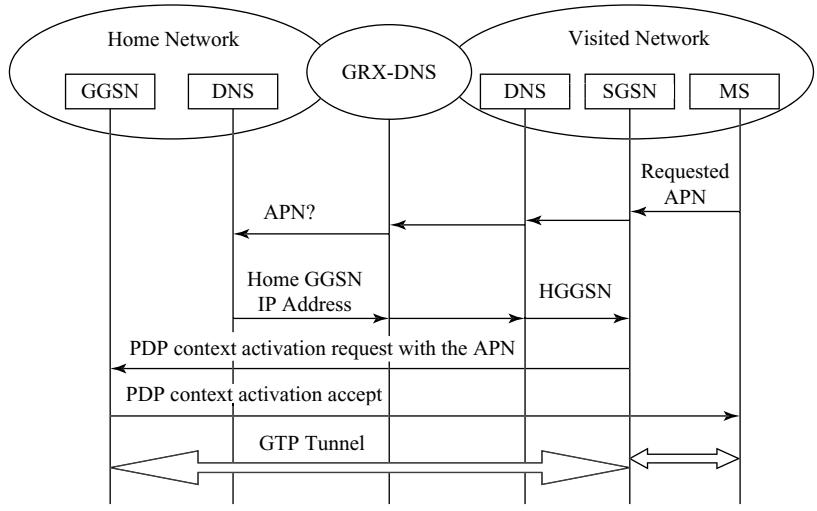


Fig. 9.16 *Example of GRX data-transfer procedure during roaming*

9.7 IP INTERNETWORKING MODEL

Internetworking is required when a user in one PLMN (say home) wants to communicate with another under a different PLMN, establishing an end-to-end communication link. The GGSN can have a number of access points to different external Packet Data Networks (PDNs). The GPRS supports internetworking with other networks based on the Internet Protocol (IP), IPv4 or IPv6. These networks may be intranet or Internet as shown in Fig. 9.17.

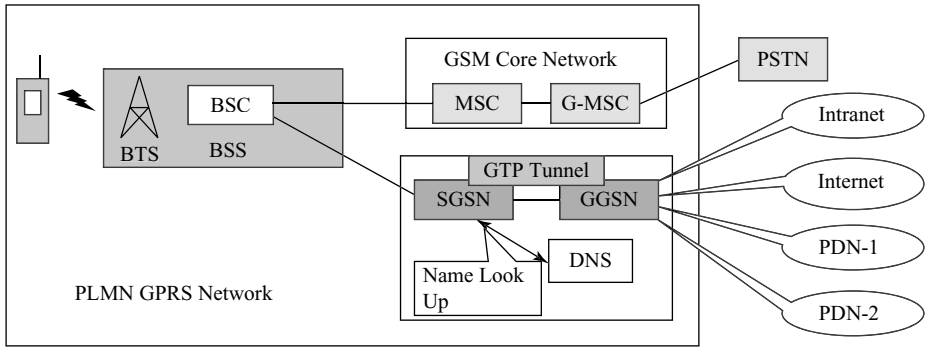


Fig 9.17 *Internetworking scenario*

The IP internetworking point is connected at G_i interface of the GPRS network. The GGSN is the entity to establish connection to the eternal PDN with the GPRS network. It is simply viewed as the IP router to other IP networks. The access to the Internet or intranet or any other PDN or Internet Service Provider (ISP) may require a user's authentication and end-to-end encryption between the MS and Internet/intranet and the allocation of dynamic address belonging to the PLMN/intranet/ISP address-

ing space. To do this, the GPRS PLMN may offer transparent or non-transparent mode of service. If the mobile user gets the IP address directly from the GGSN, it is called *transparent mode*. This address is a public IP address given either at subscription (static case) or is obtained at PDP context activation (dynamic). The transparent case provides at least a basic ISP service. On the other hand, if the GGSN asks an external Dynamic Host Configuration Protocol (DHCP) or Remote Authentication Dial-In-User (RADIUS) server from the subscriber home intranet for an IP address for the mobile device then it is called *non-transparent mode*.

The process of PDP context activation results in the mobile device obtaining an IP address. This IP address may be the operator's network or any external network. To get the IP address, a user may select the specific menu in the mobile device, for example, home PLMN or Internet. The user request is passed through the SGSN to a correct GGSN. To find the correct GGSN, the SGSN uses the Domain Name Server (DNS) within the operator's private networks. The DNS server will return the IP address of a specific GGSN where the particular connection is located. Once the PDP context is activated, the user can then use the services provided by the access point, for example, if the user is connected to Internet, he/she can surf the Web. An example of an IP inter-networking model is given in Fig. 9.18.

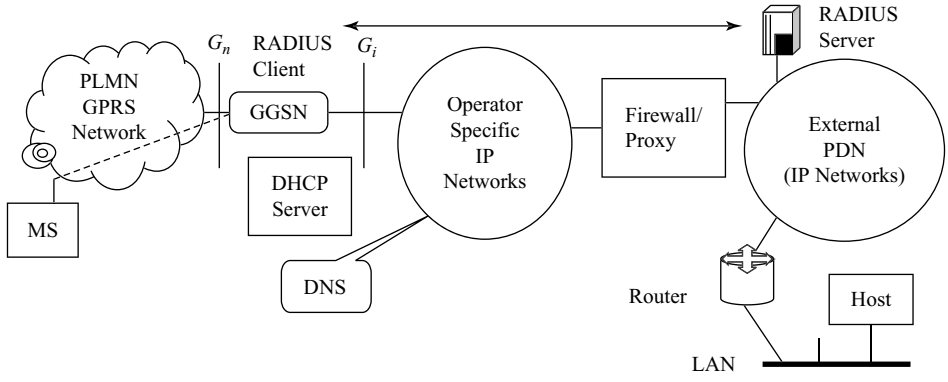


Fig. 9.18 *IP Internetworking model*

9.8 GPRS INTERFACES AND RELATED PROTOCOLS

When two persons are communicating with each other, it is considered that they understand the language with which they communicate. The actual processing is done inside the human brain. In any type of communication we may organize the whole process as a layered structure. Each peer layer must understand each other, i.e., they should communicate with common language. Information flows horizontally between the layers among the peers, flowing from one layer, down through the other layers, across the transmission medium and up through the layers on the receiving side as in the Internet model. The logical communication between two peer layers in different systems is called *external communication* as shown in Fig. 9.19, and the type of external communication is defined as the **protocols** that define every aspect of the exchange of information between two peer layers. Information may be the control data or user data. Information flow is from higher to lower layers at the transmission site and from lower to upper layer at the receiving site. This flow of information is known as direct communication. The IP packets, which are transported through the GPRS network from a mobile user to any external IP network, are encapsulated in order to transfer data across different interfaces within the network architecture. The set of protocols are defined and needed inside the GPRS network across the different entities and are discussed in this section.

The GPRS protocol architecture is organized into two planes—the *user plane*, also called the *transmission plane*, and the *control plane*, also called the *signaling plane*. To realize the transfer between the correct network nodes with required performance characteristics, the user plane protocols need information about

addresses of peer entities, specified protocols, etc. Figure 9.20 shows the protocol stacks for a user plane that are responsible for data transmission. The GPRS control plane (Fig. 9.25) realizes Session Management (SM), GPRS Mobility Management (GMM) and Quality of Service (QoS).

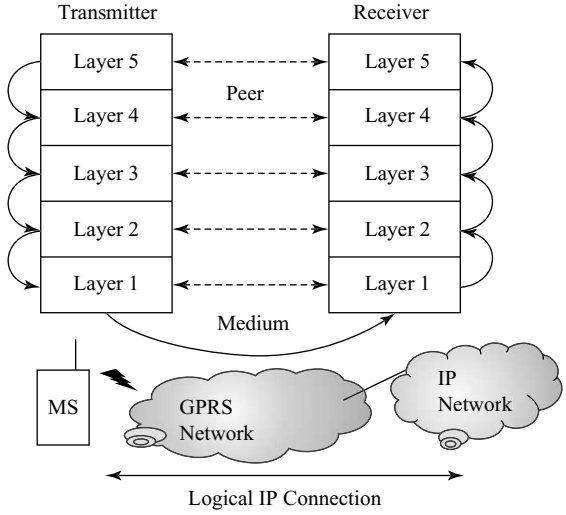


Fig. 9.19 *Layered architecture to transport data*

9.8.1 GPRS Transmission Plane

The complete overview of GPRS transmission plane protocols is given in Fig. 9.20. The application layer data is to be transferred from the MS to any external PDN over the GPRS network through the G_i interface. The MS gets a temporary IP address from the GGSN or from the PDN for IP routing. The GPRS core network interface G_n between the SGSN and GGSN is also based on IP. The IP packets are segmented and compressed by an SNDCP (Subnetwork Dependent Convergence Protocol) layer between the MS and the SGSN. Logical Link Control (LLC) is also defined between them. The G_b interface is based on frame relay protocol for handling packet data. Radio Link Control (RLC) and Medium Access Control (MAC) are the part of radio interfaces for segmenting the information into "Radio Blocks" for sharing the radio resources. In the following sections, each protocol will be explained with their major functionalities briefly.

U_m and A_{bis} Interfaces These interfaces work between the MS and the PCU. Data Link Control (DLC) at the mobile U_m interface is divided into two layers: RLC (Radio Link Control) and MAC (Medium Access Control). The RLC provides reliable logical connection between the MS and BSS, while the MAC controls the access to the physical medium and the radio link. RLC segments the LLC packets into smaller packets called 'radio block' to be transmitted over the radio link. It reassembles the radio blocks coming from A_{bis} interface into LLC packets. RLC also provides the error correction functionality and enables selected retransmission of radio blocks using ARQ (Automatic Repeat Request) protocol in case of acknowledged mode of transmission.

Radio resource is shared among multiple users using time division multiple access (TDMA) through MAC protocols. Several subscribers for GPRS packet switching are multiplexed on the U_m interface and share one allocated time slot. LLC provides a highly reliable and ciphered link between MS and SGSN. Major functionalities include sequence control, error detection and correction, flow control and ciphering. Two different logical link identifiers: TLLI (Temporary Logical Link Identifier) and DLCI (Data Link Connection Identifiers) exist. TLLI is assigned to the MS when it attaches to the GPRS network. There is one-to-one correspondence between the TLLI and IMSI of the MS that is only known in the MS and the SGSN.

GSM RF, the physical layer of the radio interface, describes the physical transmission of digital information. Error detection and correction is possible at this layer by adding checksum with the actual digital information. E1 is basically the 2048 kbps link at the A_{bis} layer.

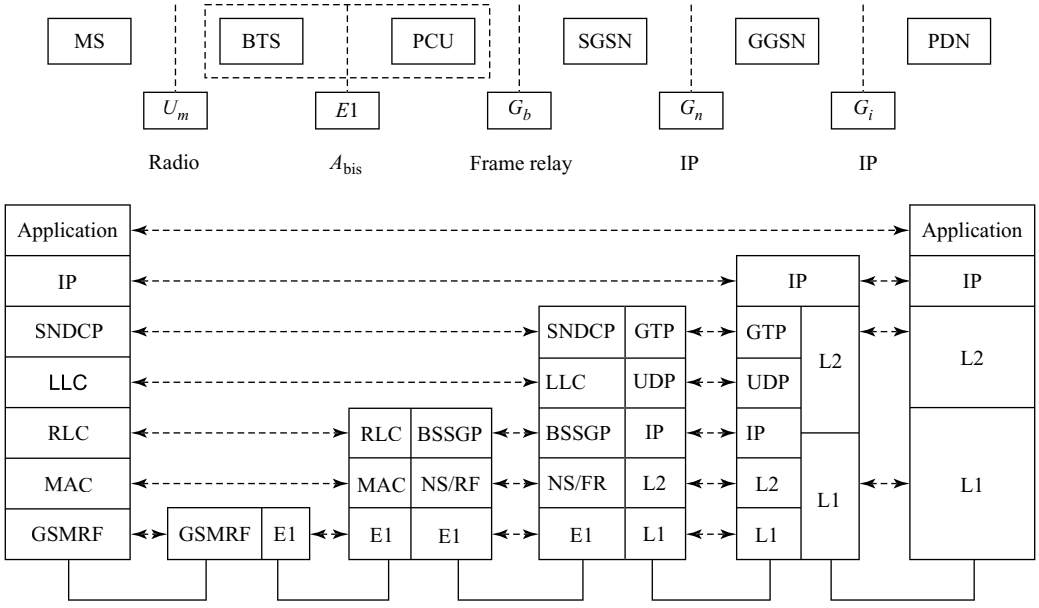


Fig. 9.20 *GPRS transmission (user) plane and protocols*

G_b Interface: Base Station Subsystem GPRS Protocol (BSSGP) The BSSGP works on the G_b interface and provides connectionless link between BSS (PCU) and SGSN. The BSSGP provides unreliable transport of LLC data units between the PCU and the SGSN and flow control in the downlink direction that prevents flooding of data in the PCU. Subscriber relevant information such as TLLI, QoS parameters, routing area, etc., are transported over the G_b interface by the BSSGP. G_b is based on the frame relay protocol in order to provide end-to-end communication between BSS and SGSN.

In downlink, each BSSGP data unit carries an LLC data unit, the identity of the MS, and set of radio related parameters. The BSSGP Virtual Connection (BVC) Identifier (BVCI) specifies the identity of the target cell. The BSS maintains one queue for each BVCI that contains the LLC PDUs for that particular cell. The SGSN is regularly informed by the BSS about the maximum queue size available for each BVC and MS at a given time. Thus, the SGSN is enabled to estimate maximum allowable data flow per BVC as well as per MS belonging to the BVC. Figure 9.21 illustrates the BSSGP.

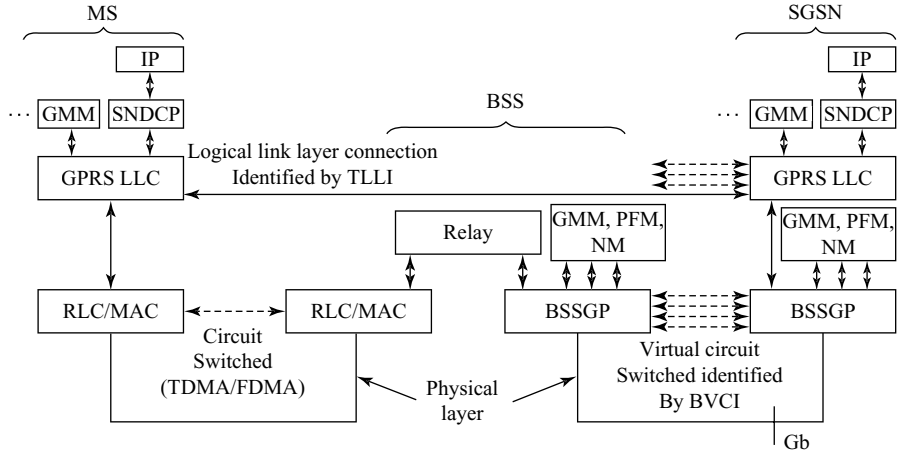


Fig. 9.21 *Illustration of BSSGP and SNDCP protocols*

SubNetwork Dependent Converge Protocol (SNDCP) When data packets are transmitted between the MS and the SGSN, the SNDCP layer compresses and segments the IP packets; and in the reverse direction, the SNDCP reassembles and decompresses the LLC packets to get the IP packets again. Towards the SGSN, the MS is identified by TLLI on the LLC layer and the PDP context is identified by NSAPI (Network Service Access Point Identifier) on the SNDCP layer. SNDCP provides adaptation of different network layers to the LLC by using NSAPI to map different PDPs onto the services provided by the LLC layer. The N-PDUs (network layer PDUs) are stored in the SNDCP layer before they are compressed, segmented and transmitted to the LLC layer. Both acknowledged and unacknowledged modes of transmission are possible on the SNDCP layer. After the addition of the SNDCP header, the compressed N-PDU is segmented to fit the maximum payload into the LLC layer.

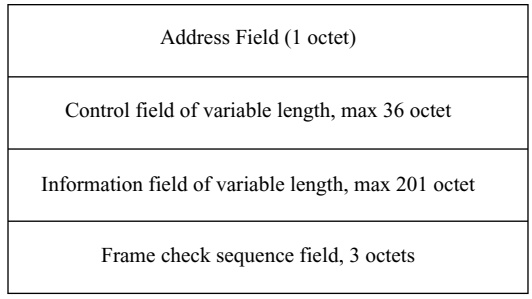


Fig. 9.22 *Frame format for LLC*

The LLC frame format is given in Fig. 9.22. LLC frames are segmented into RLC blocks. When a complete frame is successfully transferred across an RLC layer, it is forwarded to peer LLC entity.

The pictorial representation of data transmission over GPRS network is given in Fig. 9.23.

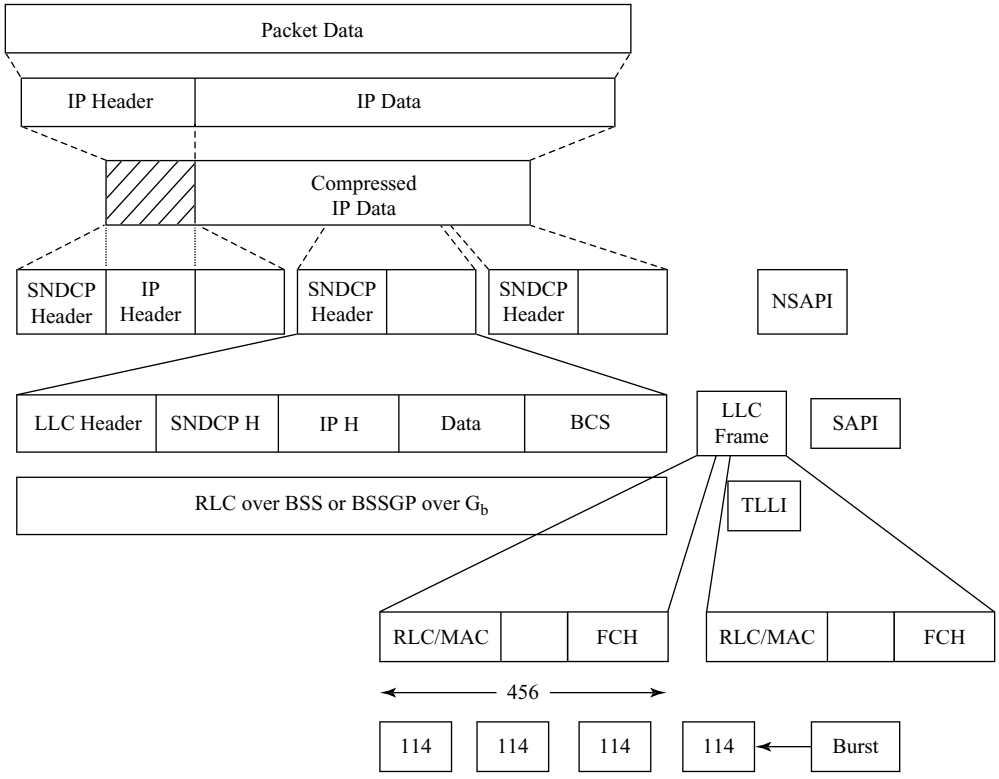


Fig. 9.23 *Network data transformation*

G_n and G_i Interfaces and GTP Tunneling Protocol Internetworking between the GPRS and other external IP networks or X.25 networks is done through the G_i interface. But, the data packets are transmitted

on the G_n interface using GTP between the SGSN and GGSN. GPRS mobility management (GMM) and Session Management (SM) are the two functionalities that use GTP signaling to create, modify and delete GTP tunnels. UDP (User Data Protocol) is the underlying protocol for GTP. Encapsulated PDUs are tunneled in between SGSN and GGSN. TID, the tunnel identifier, is used in the GTP header to indicate the particular datagram under the GTP tunnel. TID consists of the mobile country code, mobile network code, Mobile Subscriber identity (IMSI) and N-SAPI, Network Layer Service Access Point Identity. Thus, for a specified PDP context, data from any external PDN injected to the GGSN is tunneled correctly to the destined SGSN by using TID. The SGSN then delivers data to the MS. Thus, mobility related function is supported completely by GTP. There are two parts of GTP; GTP-U that is responsible for data tunneling and GTP-C, responsible for signaling between SGSN and GGSN. Figure 9.24 illustrates the GTP tunneling over GPRS networks.

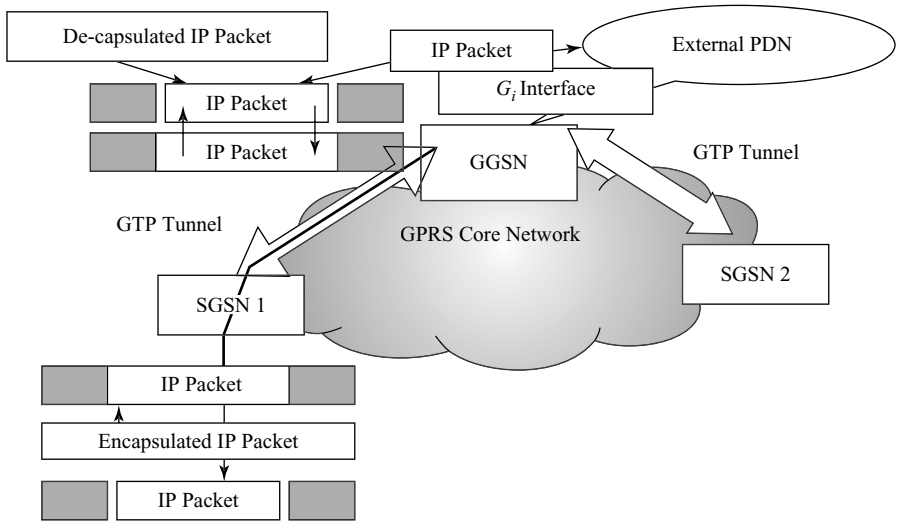


Fig. 9.24 *GTP Tunneling over a GPRS network*

9.8.2 GPRS Control (Signaling) Plane

In Fig. 9.25, GPRS control plane is shown in between the MS-SGSN and SGSN-GGSN. The GMM and SM signaling layers use Layer 2 protocols for message transfer over air interface as described earlier. The control plane between the SGSN and the registers HLR, EIR and between the SGSN-MSC/VLR is based on Signaling System number 7 (SS7) like the GSM core network. Basically, it is an enhanced version of SS7 and is known as Mobile Application Part (MAP). In GPRS, BSSAP is called BSSAP+ as it is the enhanced version of that used in the GSM network. The following interfaces work across the different nodes of GPRS control plane:

1. G_c – between the GGSN and HLR/GR
2. G_f – between SGSN and EIR
3. G_r – between SGSN and HLR/GR
4. G_s – between the SGSN/SLR and MSC/VLR
5. G_d – between SGSN and MSC connected to SMS centre for short message service

The SS7 is a very important signaling protocol that is used in ISDN (Integrated Services Digital Network) and GSM PLMN. It is an out-of-band signaling and prerequisite for ISDN. In the GPRS network, elements communicate with MSC, VLR and HLR, so the SS7 protocol has to be implemented on the interfaces among which entities are communicating. The lowest three levels of the SS7 protocol is MTP 1,2,3 (Message

Transfer Part) and corresponds to layers 1, 2 and 3 of the OSI model. The main functions of MTP are secure message transfer, routing and physical transmission of bits. MTP 1 is physical layer protocol. The information is physically transported on E1 link with 2048 kbps. MTP 2 is for error correction and MTP 3 is for addressing.

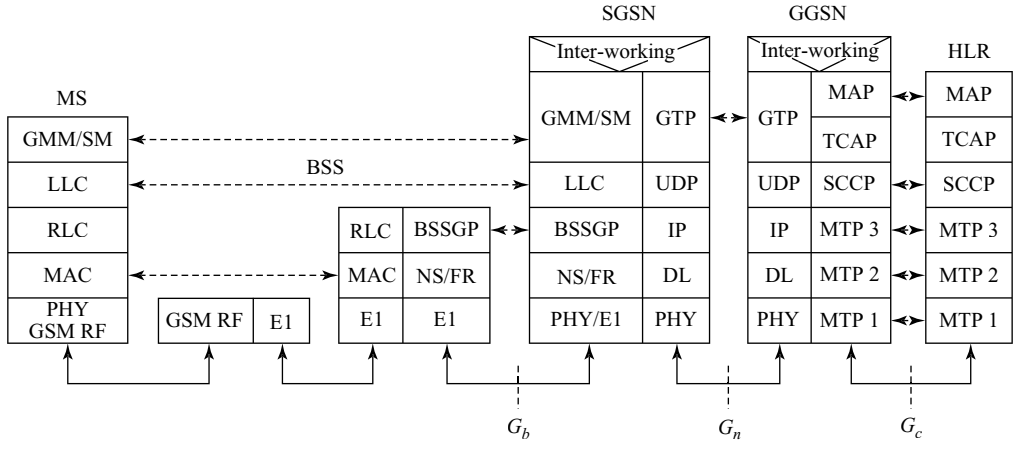


Fig. 9.25 *GPRS signaling plane*

SCCP (Signaling Connection Control Part) enables the use of global titles for transfer of signaling information without a relationship to a user channel.

The SGSN application part is used on G_r interface with the MAP HLR in the HLR. It allows database enquiries for mobility managements.

The BSSAP+ is deployed on the G_s interface to support those MS which are used both for GSM and GPRS network. Since SS7 is deployed on the G_r interface, a part of the whole procedure is performed on SS7. The message and their parameters vary depending on the user or application part exchanging information.

The signaling information between the SGSN and HLR is transported through the G_r interface. G_f is an optional interface implemented in between SGSN and EIR. The common mobility management (CMM) functionality is maintained through the G_s interface in between SGSN and MSC/VLR. The G_c interface between GGSN and HLR, necessary for incoming data application, is based on MAP. The G_d interface between SGSN and MSC is connected to the SMS server and is necessary for transporting SMS messages. Figure 9.26 shows the different interfaces used in GPRS control plane.

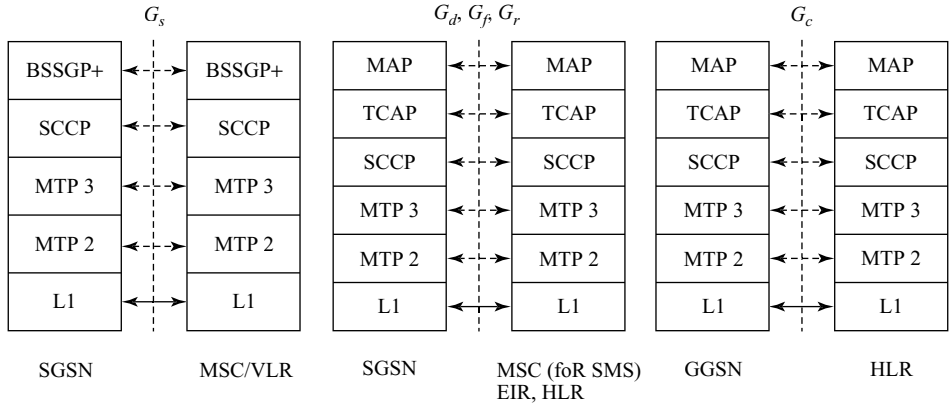


Fig. 9.26 *GPRS control plane interfaces and protocols*

9.9 GPRS APPLICATIONS

The tremendous growth of mobile subscribers during the last decade has necessitated the use of Wireless Application Protocol (WAP) to be interfaced with the Internet services. The WAP protocol suite enables network operators to offer many Internet services optimised for mobile use. These may be Internet applications such as news and weather forecasts; entertainment services like games, multimedia messages and video streaming applications, or business applications within the workplace such as accessing email services.
A wide range of corporate and consumer applications are enabled for non-voice mobile services such as SMS and GPRS data packet services. Some of the applications will be discussed briefly in this part.

Chat Similar to Internet chat groups that have proven a very popular application of the Internet, groups of likeminded people having common interests can use non-voice mobile services as a means to chat, communicate and discuss. Because of the synergy of the Internet, GPRS allows mobile users to participate fully in existing Internet chat groups rather than set up their own groups dedicated to mobile users. The number of participants is an important factor determining the value of the news group where use of GPRS would be advantageous. SMS would remain the primary bearer for chat applications in the future.

Textual and Visual Information A wide range of content can be delivered to mobile phone users with information for share prices, sports scores, weather, flight information, news headlines, jokes, traffic conditions or location-based services. This information may not only be text but graphical or other types of visual information transport is also possible. Quantitative information like share prices, sports scores or daily temperature, can be transported through 160-character SMS service. However, information like news stories cannot be sent through 160 characters. As such, GPRS will be used for qualitative information transport when end users have GPRS capable devices, but SMS will continue to deliver quantitative services.

Still and Moving Images Photographs, greetings cards, and static web pages can be sent and received over mobile networks. Like real-time desktop publishing it will be possible with GPRS to post images from a digital camera connected to a GPRS radio device directly to an Internet site. Sending moving images in a mobile environment has several vertical market applications like monitoring parking lots and building sites against intruders, or for telemedicine and video conferencing.

Web Browsing Mobile Internet browsing would be better suited for Web browsing than circuit-switched data because of slow speed of circuit-switched connections.

Document Sharing for Collaborative Work People in different places working on the same document at the same time can share data through mobile devices. This kind of application could be useful in solving problems related fire fighting, medical treatment, and news updates. Multimedia applications consisting of text, pictures, voice and video can be envisaged in such cases. By providing sufficient bandwidth, GPRS facilitates multimedia applications such as document sharing in collaborative work.

Corporate Email Within a corporate section, if a large number of employees have to be away from their desks at office hours, it would be possible to keep them up-to-date by extending the corporate email systems beyond the employee's office work station. GPRS capable devices are expected to be widespread in corporate sections. So, corporate email applications using GPRS rather than the Internet email may have a huge market potential.

File Transfer File transfer applications encompass any form of downloading across a mobile network. This could be a presentation document for a traveling worker, an appliance manual for a service engineer or any software application. The source of this information could be one of the Internet communication methods such as FTP, telnet, and http. Irrespective of the source and type of files being transferred, this type of application is bandwidth demanding. It may therefore require a high-speed mobile data service such as GPRS, EDGE or UMTS to run satisfactorily across a mobile network.

Financial Applications These could be banking services such as payment procedures via a mobile account, online transaction and access to bank accounts.

Location Dependent Services This includes traffic condition on roads, accident information, hotel information or any information conveyed via the mobile and related to a particular location of user interest.

Summary

GPRS is a new bearer service for GSM that greatly improves and simplifies wireless access to packet data networks, e.g., to the Internet. It applies a packet radio principle to transfer user data packets in an efficient way between mobile stations and external packet data networks. Users of GPRS benefit from shorter access times and higher data rates. In conventional GSM, the connection set-up takes several seconds and rates for data transmission are restricted to 9.6 kbps. GPRS in practice offers session establishment times below one second and ISDN-like data rates up to several tens of kbps. In addition, GPRS packet transmission offers a more user-friendly billing than that offered by circuit-switched services. In circuit-switched services, billing is based on the duration of the connection. This is unsuitable for applications with bursty traffic. The user must pay for the entire airtime, even for idle periods when no packets are sent (for example, the user reads a Web page). In contrast to this, with packet switched services, billing can be based on the amount of transmitted data. The advantage for the user is that he or she can be 'online' over a long period of time but will be billed based on the transmitted data volume. This discussion apparently shows that the GPRS system is a versatile and cost-effective solution for the wireless packet-data connectivity supporting global roaming.

References

[1] Stuckmann, Peter, *The GSM Evolution–Mobile Packet Data Services*, John Wiley, 2003.

[2] Bettstetter, C. H.J. Vogel, and J. Eberspacher, *GSM Phase 2+ General Packet Radio Service GPRS: Architecture, Protocols, and Air Interface*, IEEE Communication Surveys, Vol 2, No.3, 1999.

[3] Bannister, J. P. Mather, and S. Coope, *Convergence Technologies for 3G Networks*, *IP,UMTS, EGPRS and ATM*, John Wiley, 2004.

[4] Sanders, G. L. Thorens, M. Reisky, O. Rulik, and S. Deylitz, *GPRS Networks*, John Wiley, 2003.

[5] Brasche G. and B. Walke, *Concepts Services and Protocols of the new GSM Phase* 2+ *General Packet Radio Service*, IEEE Communication Magazine, pp.94–106, August 1997.

[6] 3GPP, *General Packet Radio Service (GPRS); Base Station System (BSS)-Serving GPRS Support Node (SGSN); BSS GPRS Protocol (BSSGP)*, Technical Specification 3GPP TS 08.18, version 6.8.0, Release 1997, June 2001.

[7] 3GPP TSG CN, *General Packet Radio Service (GPRS); GPRS Tunneling Protocol (GTP) across Gn and Gp interface (3GPP TS 29.060 version 5.0.1 Release 5)*. Technical Specification, 3rd Generation Partnership Project, Jan 2002.

[8] 3GPP TSG CN, *General Packet Radio Service (GPRS); Interworking between the Public Land Mobile Network (PLMN) supporting GPRS and Packet Data Networks (3GPP TS 29.061 version 5.0.1 Release 5)*. Technical Specification, 3rd Generation Partnership Project, Jan 2002.

[9] Blyth, K.J, and A.R. Cook, *Designing a GPRS Roaming Exchange Service*. In Proc. of the Second International Conference on 3G Mobile Communication Technologies, London, UK, March 2001.

Questions for Self-Test

9.1 The role of SGSN is like a switching centre of MSC in GSM networks.
a. True b. False

9.2 For packet data service, the location register of SGSN stores information of the
a. current cell b. user profile
c. IMSI addresses for all packet data users d. all of above

9.3 In GPRS packet data services, mobility is managed by
a. MSC b. SGSN c. GGSN

9.4 GPRS network basically integrates a packet-based air interface into an existing circuit-switched GSM network.
a. False b. True

9.5 The interfacing node towards external PDN or the other PLMN is
a. SGSN b. GGSN c. GMSC

9.6 Border Gateway is an IP router that connects different operators.
a. False b. True

9.7 To get GPRS services, mobile stations need to register with
a. SGSN b. GGSN c. MSC

9.8 Packet Temporary Mobile Subscriber Identity (P-TMSI) is assigned to the mobile during
a. GPRS attach procedure b. GPRS detach procedure

9.9 An important task of DNS is the address resolution for GGSN IP address during the process of PDP context activation.
a. True b. False

9.10 BSSGP protocol provides unreliable transport of LLC data units between the
a. PCU and SGSN b. BSS and MS

9.11 The underlying protocol for GTP is the UDP.
a. True b. False

9.12 GRX takes care of billing in the roaming scenario.
a. True b. False

9.13 Border router is placed across different IP networks to connect with the backbone network.
a. False b. True

9.14 GPRS uses the same frequency standard allocated as in GSM.
a. True b. False

9.15 Subscriber information from HLR to GGSN is passed through _____ interface.

9.16 _____ switching is the major advantages of GPRS network?

9.17 The components of TID (tunnel identifier) are
a. _____
b. _____
c. _____
d. _____

9.18 The Gb interface is based on _____.

9.19 In the uplink, the SNDCP _____ and _____ IP packets.

9.20 In the downlink, the SNDCP _____ and _____ the LLC packets into IP packets.
9.21 BSSGP works between BSS and _____
9.22 E1 is the physical link of _____kbps at the A_{bis} interface.
9.23 Getting IP address directly from GGSN is called _____ access.
9.24 There are two planes in GPRS architecture: _____ and _____.
9.25 The main functions of control plane are _____, _____ and QoS.
9.26 GPRS is the extension of GSM network. Explain.
9.27 How are channels allocated for voice and data in GPRS network?
9.28 What are the main classification of user equipments in GPRS network?
9.29 Describe the different types of GPRS services available.
9.30 What is the role of GRX (GPRS exchange)?
9.31 Define non-transparent mode of getting IP address in a GPRS network.
9.32 What are the main functionalities of RLC and LLC protocols in the GPRS network?
9.33 What is the role of SGSN in the GPRS networks?
9.34 What are the functionalities of GGSN node in the GPRS networks?
9.35 Explain the PDP context activation in GPRS network?
9.36 How is data transferred to external PDN through GPRS network?
9.37 Describe the role of PDP and GTP tunneling protocol in a GPRS network in connection to packet data transportation.
9.38 What are the major functionalities of BSSGP and SNDCP protocols?
9.39 Give the pictorial representation of GPRS user plane and control plane protocols and their working.
9.40 What are the main interfaces involved in relation to signaling plane of GPRS networks?
9.41 Highlight some important applications in GPRS networks.

Overview of CDMA-Based IS-95 2G Cellular Networks

10

Introduction

Code Division Multiple Access (CDMA) was created to provide secure communication and navigation systems for military applications. It requires the development of spread-spectrum technology for multiple accesses over a single carrier frequency and reduction of interference. The advantage of CDMA for personal communication services is its ability to accommodate many users on the same frequency at the same time. Interim Standard 95 (IS-95) based on CDMA also known as cdmaone is a popular 2G CDMA cellular network widely deployed in North America, Korea, Japan, China, South America and Australia. There are many advantages of CDMA over TDMA and FDMA, one of which is the increased capacity. The IS-95 system was developed so that it could be compatible with the US analog cellular system AMPS frequency bands. So, dual-mode operation of mobile handsets and base stations were needed. CDMA/AMPS dual-mode phones were first produced by Qualcomm in the year 1994. IS-95 can support 64 users, which are orthogonally coded and simultaneously transmitted on each 1.25 MHz channel. It is based on the direct-sequence spread spectrum technology (DS-CDMA).

In this chapter, the historical development of CDMA technology along with an overview of IS-95 cellular network standards will be provided.

10.1 CDMA EVOLUTION

At the beginning of the evolution process in 1G cellular systems, the quality of voice transmission and more mobility of users were the primary focus. Later, capacity enhancement became an important issue in the evolution process of 2G digital systems such as CDMA and TDMA in North America and GSM in Europe.

The Cellular Telecommunications and Internet Association (CTIA) released the user performance requirements for 2G cellular systems in September 1998 with requirements of a tenfold increase in capacity with respect to AMPS (Advanced Mobile Phone System, analog system), and privacy for voice and data communications along with the ability to support new services. It also considered the compatibility of the existing analog frequency spectrum, reasonable infrastructure and mobile terminal costs.

In 1990, Qualcomm developed and demonstrated a CDMA-based digital cellular system that claimed a twenty-fold increase in capacity over analog systems. In January 1990, the Telecommunication Industry Association (TIA) began the standardization process and in March 1990, the subcommittee on digital cellular systems IS-95 TR 45.5 was created for developing IS-95.

In 1992, following the instruction of CTIA, the TIA TR47.5 initiated standardization work on wideband spread spectrum technologies for cellular applications developed by Qualcomm with the support of other companies like Lucent, Motorola, and Nortel.

CDMA IS-95 belongs to the second generation of cellular systems. In July 1993, the first CDMA standard IS-95 was published. It defines the compatibility requirements for 800 MHz AMPS and CDMA systems. The CDMA Development Group (CDG) was established in 1994 with the mission to promote IS-97. The IS-95 standard for the CDMA common air interface was adopted in 1993 followed by an enhanced and revised version (IS-95A) in 1997. The operational features for IS-95A include a wideband channel bandwidth of

1.25 MHz, chip rate of 1.2288 Mcps (chips per sec), 14.4 kbps for voice communication, provision for power control and call processing and handoff. Commercial operation of CDMA IS-95 systems started in the year 1996.

The IS-95B standard was released in 1998 with a data rate capacity of 117.2 kbps combining 8 Walsh codes. It was implemented in Korea, Japan and Peru.

Several standardisation processes began with the idea of increasing the data rate and user connectivity around the world during the period 1995–1998, and the new third generation (3G) systems based on CDMA was developed.

In Europe, FMA2 [FRAMES (Future Radio Wideband Multiple Access System) Multiple Access based in wideband CDMA] gave birth to the WCDMA (Wideband CDMA) standards later known as UMTS (Universal Mobile Telecommunication Systems). In US, the TR 45.5 group was involved in the 3G-standardization process of cdma2000 systems originally known as IS-95C.

The Third-Generation Partnership Project (3GPP) was created to produce a common standard for asynchronous wideband systems bringing together several standard organizations (ARIB and TTC in Japan, ETSI in Europe, T1P1 in USA, TTA in Korea and CWTS in China) under the 3GPP umbrella. They developed a standard that was backward compatible with the existing GSM/GPRS network structure.

The Third-Generation Partnership Project 2 (3GPP2) was created to harmonise the use of the multi-carrier CDMA2000, initially consisting of two TR 45.5 (TIS-USA) and TTA (Korea) groups, later joined to ARIB, TTC and CWTS. Ultimately, the harmonization process concluded and generated three operating concepts based on the following:

1. Direct Sequence (DS) spread spectrum used in UMTS in Europe.
2. Multi Carrier (MC) multiple access used in cdma2000 prevalent in USA.
3. Time Division Duplex (TDD) operation in time, an UMTS operation presently preferred by China.

The cdma 2000 represents a family of technologies as underlined:

1. **CDMA2000–1XRTT:** 1X represents a single carrier and RTT stands for Radio Transmission Technology.
2. **CDMA2000-1XEVDO:** EV stands for evolution and DO for data optimized.
3. **CDMA2000-1XEVDV:** DV stands for data and voice.
4. **CDMA2000-3X RTT:** 3X represents the use of multi-carriers.

Figure 10.1 illustrates the evolution path of CDMA systems.

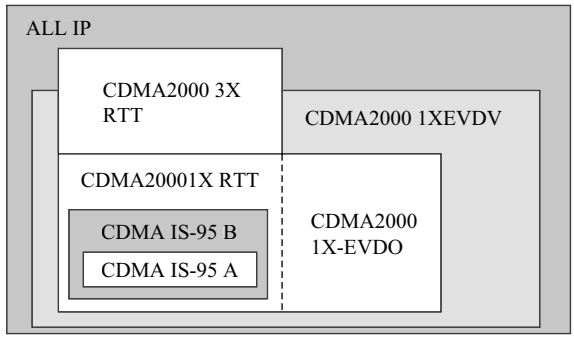


Fig. 10.1 *Evolution for CDMA systems*

10.2 CDMA IS-95 SYSTEMS

The CDMA IS-95 system is the second-generation digital cellular network system. In this section, the basics of the IS-95 system and link characteristics are given.

CDMA networks usually employ full-duplex access methods, such as Frequency Division Duplex (FDD). The carrier transmits forward link channels or downlinks from base transceiver stations to Mobile Stations (MSs), whereas another frequency is allocated for the reverse link or uplink, transmitting from MSs to their server BTSs. A BTS can utilize more than one pair of forward and reverse carrier frequencies to increase capacity in its cell. Carriers have 1.23 MHz of bandwidth but the separation between carriers is 1.25 MHz, as shown in Fig. 10.2. Thus it is common to see the CDMA carrier width declared both as 1.23 MHz and

1.25 MHz. Separation between uplink and downlink carriers is usually 45 MHz for an 850-MHz band and 80 MHz for the 1.9-GHz band. In the IS-95 system, 869–894 MHz is used for forward link and an 824–849 MHz band is used for reverse link.

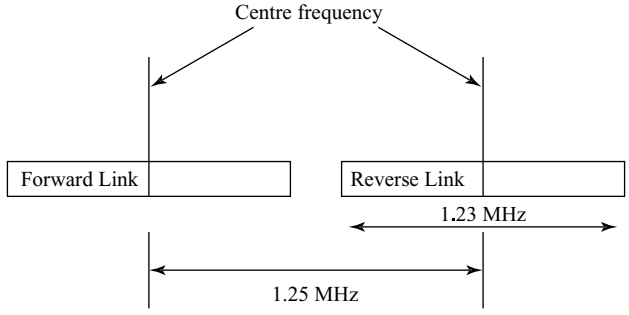


Fig. 10.2 *IS-95 carrier*

The IS-95 allows each user within a cell to use the same radio channel. Also, the users in an adjacent cell use the same channel because of DS-spread spectrum technology, thus eliminating the need of frequency planning.

Depending on the channel condition and other user activities, the user data rate in IS-95 may vary in a real-time system. IS-95 uses different modulation and spreading techniques for the forward and reverse links. The maximum user data rate is 9.6 kbps. User data in the IS-95 is spread to a channel chip rate of 1.2288 MCPS using a combination technique. The spreading process is different for forward and reverse links in the original CDMA specification. Every CDMA channel in any BTS is identified by two parameters—an RF carrier (radio frequency) and a code. The first parameter defines the 1.25 MHz carrier center frequency, whereas the second consists of a Walsh code or an offset mask for Pseudorandom Noise (PN) long code sequence. In the forward link, the user data stream is encoded using a ½ rate convolutional code, interleaved and spread by one of 64 orthogonal sequences (Walsh code). Each mobile in a cell is assigned a single spreading code, providing perfect separation among the signals from different users. Orthogonality among all users in the forward channel is maintained by scrambling the signal using a PN sequence having a length of 2^{15} chips. A pilot channel (code) is transmitted in a forward channel at a higher power than the user channel. By this pilot channel, the user can determine and react to the channel condition while employing coherent detection.

A different spreading technique is required at the reverse channel because signals from different users arrive with different propagation path at the base station. The data stream in a reverse channel is first convolutionally encoded with a 1/3 code rate. After interleaving, each block of six encoded symbols is mapped to one of the 64 orthogonal Walsh functions, providing a 64-ary orthogonal signaling.

Tight power control is used in the base station for each user so as to avoid near–far problem that arises from varying received powers by the users. At mobile and base stations, the RAKE receiver is used to resolve and combine multipath components to reduce the degree of fading.

10.2.1 Forward Link in CDMA IS-95 Systems

Four types of channels are present in a forward link. These are one pilot channel, one synchronization channel, up to seven paging channels and up to sixty-three traffic channels.

Each pilot transmits the same spreading sequence at a different time offset, which can be used to distinguish signals of different pilots. Every CDMA IS-95 carrier has a Forward Pilot Channel (FPiCh) that uses Walsh code W_0^{64}. A mobile station acquires timing synchronization for the forward link through this

pilot channel. It also provides a phase reference for coherent demodulation, and compares signal strengths between base stations during handoff. The Synchronization Channel (SyncCh) broadcasts synchronization messages to the mobile stations. After acquisition and successful processing of the SyncCh, the MSs can identify their server base station. SyncCh uses Walsh code W_{32}^{64}. A unique offset number, ranging from 0 to 511, does the identification of serving the base station.

Through the Forward Paging Channel (FPCh) control, the base station to the mobiles broadcasts information and paging messages. It uses Walsh code W_1^{64} to W_7^{64}, where the first code is the default carrier for primary forward paging channel. The base station sends the identification of CDMA carriers to be used by the mobiles in the forward and reverse channels. All required information in accessing the CDMA network is transmitted through this FPChs. Typical messages on the paging channel include pages, traffic channel assignments and short messages.

The Forward Traffic Channels (FTChs) are the logical channels used for call handling. They use Walsh code W_8^{64} to W_{63}^{64} except the 32, which is used for SyncCh. Figure 10.3 illustrates the IS-95 forward channel. Mobiles continuously monitor pilot channels while on a paging or traffic channel so that they can move to a different cell.

Data in the forward traffic channel is grouped into 20 ms frames. The user data is first convolutionally coded and then formatted and interleaved for the adjustment of actual user data rate. Data rates are flexible (1200, 4800 or 9600 bps) in order to support variable rate voice coders. The signal is then spread with a Walsh code and scrambled by a long PN sequence at a rate of 1.2288 Mcps.

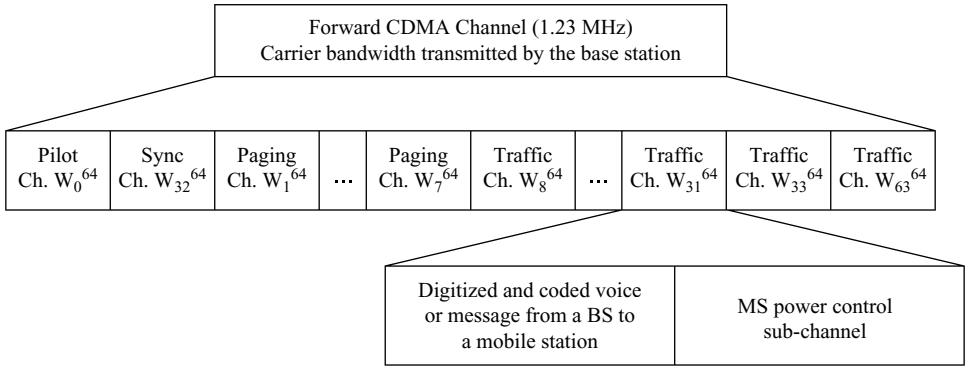


Fig. 10.3 *IS-95 forward channel*

10.2.2 Reverse Link in CDMA IS-95 Systems

The reverse channel is a composite of all the outputs from all the mobiles in the base station's coverage area. At any time, there may be *m* mobiles engaged in a call and *n* mobiles trying to gain access to the system.

In the reverse link there are two logical channel types—Reverse Access Channel (RACh) and Reverse Traffic Channel (RTCh), both sharing the same frequency assignment. Reverse channels are identified by long code-offset masks for both access or traffic channels. The reverse link is asynchronous due to the distinct signal propagation delay from the mobile stations to the base stations. The reverse channel structure allows up to 62 different traffic channels and 32 different access channels.

Communication is initiated with the base station by a mobile through the RACh. It is also used for the response of a specific message such as response to a paging channel or order sent by the base station. The access channel is a random access channel which identifies each channel by a long PN code. Each RACh is uniquely associated to a paging channel (FPCh). There are 32 RAChs per FPCh.

During a call process, the mobile station uses a Reverse Traffic Channel RTCh. It is assigned to a base station for both forward and reverse links. The channel assignment information is transmitted to the mobile over the associated FPCh. Figure 10.4 illustrates the reverse link channel in IS-95.

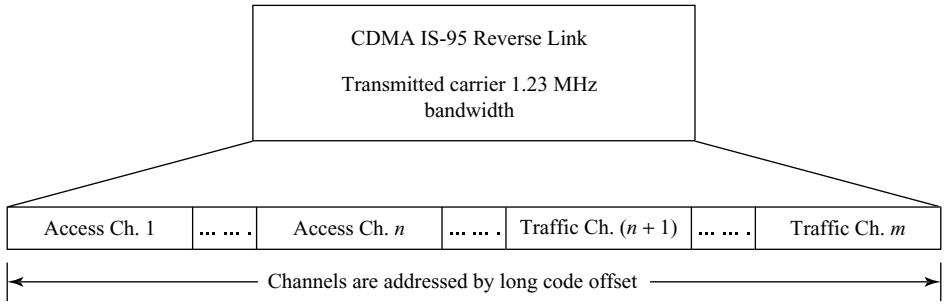


Fig. 10.4 *IS-95 reverse link channel*

10.2.3 About PN Sequences Related to CDMA IS-95

Each CDMA carrier can accommodate several logical channels modulated and multiplexed (on the reverse link) by PN sequences. In CDMA IS-95 systems, there are three types of PN sequences performing these functions:

Two short PN sequences (PN-I, PN-Q) and one long PN sequence (PNLC). All of these PN sequences in a CDMA system are synchronised with a time reference. A brief discussion about these sequences is given in this section.

Short PN Sequences, PN-I and PN-Q A circuit with 15 shift register stages generates short PN sequences in CDMA IS-95. Two distinct Maximum Length Sequences (MLSs) are generated with the following two polynomials for PN-I and PN-Q:

$$\text{PN–I:} \quad p(x) = x^{15} + x^{13} + x^9 + x^8 + x^7 + x^5 + 1 \tag{10.1}$$

$$\text{PN–Q:} \quad p(x) = x^{15} + x^{12} + x^{11} + x^{10} + x^6 + x^5 + x^4 + x^3 + 1 \tag{10.2}$$

Each of the circuit generates a 32767–chip-long sequences ($2^N - 1$), which have 16384 '1s' and 16383 '0s'. An external circuit is used to insert an extra '0' into each sequence after counting 14 consecutive '0s' to make PN sequence even having equal '1s' and '0s'. However, this occurs only once on a complete set of 32767 chips. Considering the transmission rate of these sequences to be 1.2288 Mcps and length to be 32768–chip long, the repetition rate is 35 times per second. PN-I and PN-Q sequences modulate all logical channels in the forward link and are used as the time reference for synchronization acquisition between the base and mobile stations. All base stations always use the same PN-I and PN-Q sequences. The set of 32768 chips is divided into 512 sets of smaller 64 chips, each set representing one distinct phase offset known as PN offset. This PN offset is used to match the base stations to the mobiles. One chip offset represents a distance of 244.1 m. Thus a sequence offset can be mapped into distance when considering signal propagation time. The reverse link does not use PN offset, but the locally generated PN-sequences are still to be synchronized with the initial CDMA reference time.

The most common MLS generator consists of a shift register working according to a specific logic that provides a feedback path, which combines the states of two or more stages of the shift register. The length of the MLS = $2^N - 1$ chips, where N is the number of stages used in the generator circuit. If the shift registers are not connected properly, the generated sequence will not be an MLS. Figure 10.5 illustrates the 3-stage shift register PN sequence generator for MLS.

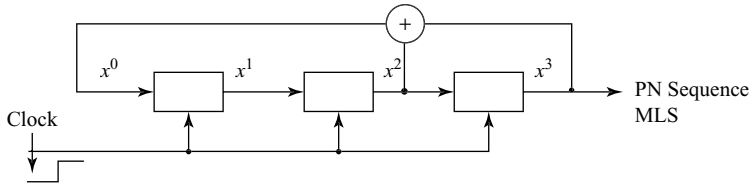


Fig. 10.5 *3-stage shift register PN sequence generator*

The polynomial for a 3-stage shift register for MLS is $p(x) = x^3 + x^2 + 1$.

Long Code PN Sequence (PNLC) The PNLC is used in the forward channel for data scrambling. The long PN sequence is generated by a 42-stage shift register circuit with MLS $(2^{42} - 1)$ chips. The polynomial used for the circuit is given by

$$P(x) = x^{42} + x^{35} + x^{33} + x^{31} + x^{27} + x^{26} + x^{25} + x^{22} + x^{21} + x^{19} + x^{18} + x^{17} + x^{16} + x^{10} + x^7 + x^6 + x^5 + x^3 + x^2 + x^1 + 1 \quad (10.3)$$

The transmission rate of the sequence is 1.2288 Mcps. The long code is 4398046511103–chip long. Repetition occurs after every 42 days, 10 hours, 12 minutes and 19.4 seconds approximately. The sequence starts with '1' and is followed by consecutive 41 '0's. All mobile stations and base stations have a long code PN generator. After the initialization process in the MS, the best-selected base stations synchronize their long codes. The PLNS is the unique code used for the entire network with a different mask of bits applied to the sequence in order to distinguish user specific characteristics.

There are two types of long code masks in CDMA standards.

1. **Public Long Code Offset Mask:** This mask varies according to the application. All CDMA calls are initiated using the public mask.
2. **Private Long Code Offset Mask:** IS-95 standard has defined the structure of this mask. The distribution of this mask is controlled by TIA and supervised by the US International Traffic and Arms Regulation (ITAR) and the Export Administration Regulations (EAR). The private offset long code provides full privacy to users.

The structure of public and private offset masks for a long code sequence applied for IS-95 is shown in Fig. 10.6.

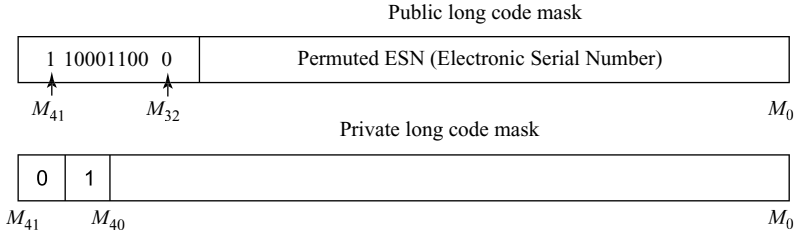


Fig. 10.6 *Long code mask format used in IS-95*

The public long code specified through M_{41} to M_{32} is a set of 1100011000; and M_{31} to M_0 is the permuted number of the mobile's Electronic Serial Number (ESN). The ESN is permuted to avoid close correlation of two mobiles that may have consecutive ESNs. The use of 32-bit ESN generates 2^{32} combinations of distinct masks. The private masks specified through M_{41} to M_{40} are a set of 01 bits, and M_{39} through M_0 are set by private procedure.

10.3 CALL PROCESSING STEPS IN CDMA IS-95 SYSTEM

Call processing consists of the exchange of messages through which a mobile station and its server base station negotiate origination and termination of calls. To establish a call, an MS must go through several states consisting of a set of commands and procedures to be executed according to the configuration parameters stored internally in the MS or transmitted through messages. The call processing steps are given as under:

Mobile Station Initialization State The MS selects the operation band and the technology of operation in case of dual band operation for AMPS and CDMA. If CDMA, the mobile station starts acquisition and synchronization (using Forward Pilot Channel (FPiCh)) of the pilot channel for this system. For acquisition of the sync channel, the MS gathers more information about the CDMA network timing and configuration parameters. The MS tunes to the SyncCh using Walsh code 32. For timing change, the MS has a small period of time to adjust the PLNC code generator and internal clock to the system. Every time the MS loses synchronisation or receives an out-of-order message confirmation, it has to restart the *mobile station initialization state*.

Mobile Station Idle State In this state, the MS monitors messages transmitted through signaling and control channels such as Forward Paging Channel (FPCH), Forward Broadcast Control Channels (FBCCh), Forward Common Control Channels (FCChs) and Quick Paging Channels (QPChs).

System Access State In this state, the MS attempts to access the system by sending messages or responding to the orders obtained from the server base stations through a Reverse Access Channel (RACh). The system access state is again divided into substates:

1. **Update overhead information substate:** The MS monitors FPChs or FCChs to update all system overhead messages.
2. **Page response substate:** The MS sends a page response message required by the server base station.
3. **Mobile station order or message response substate:** The MS sends responses related to a specific message or order previously sent by the system.
4. **Mobile station origination attempt substate:** The MS sends an origination message to the server BS.
5. **Registration access substate:** The MS sends messages to start registration.
6. **Mobile station message transmission substate:** The MS transmits data burst messages to the system.

Mobile Station Control on the Traffic Channel State The MS establishes a communication path between itself and its serving BS. On forward and reverse links, a communication link is established between an MS and a BS during a call, using fundamental channels and dedicated control channels to send and receive data messages and voice. There are several substates for this last step of call processing:

Traffic Channel Evaluation SubState The MS evaluates and monitors for the possibility of receiving data messages on any traffic channel (FTCh) and starts transmitting on the corresponding reverse channel, but some configurations must be set before transmission and reception on the forward and reverse channels. A message exchange process establishes and defines configuration between the MS and the BS.

Waiting for Order Substate The mobile terminal waits for an alert with information message.

Waiting for Mobile Station Answer Substate The MS waits for the user to answer the call or to perform some other appropriate action. CDMA systems have a timer designed specifically (T = 65 s) for occasions when the subscriber does not answer a call. If the pre-fixed time expires before answering or following any instruction (voice mail, hold), etc., the MS turns off the transmitter circuit and enters the system determination substate. In this state, the MS also monitors FTCh, transmission power adjustment according to power control procedure and handoff.

Conversation Substate The MS transmits and receives traffic data frames to/from its serving base station according to the current service configuration options. It also performs gating operation of transmitted Reverse Pilot Channel (RPiCh) depending on the power control processes: active mode or control hold mode. When in active mode, the MS transmits and the RPiCh is not gated. When in control hold operation, the MS does not transmit, but may wait for handoff process to be concluded or expect a resource allocation for supplemental channel use.

Release Substate The MS releases all resources used during the call processing, disconnects the current call and enters into the system determination substate.

10.4 POWER CONTROL

Power control is one of the very important operational requirements in CDMA networks. Power control is a must for good quality of signal reception and increasing capacity because an MS generates interference for all other users within the same base station under the coverage area and also for the neighbor cells. As the number of users is ever increasing, for increased numbers of active users, the interference level also increase. CDMA systems implement power control techniques to minimize interference by lowering the MS transmitting power, the interference level and increasing the system capacity.

For power control, the system analyzes network quality parameters. The most important parameter is the Frame Bit Error Rate (FER). Based on FER, power control quickly estimates the minimum MS transmission power to achieve the required communication quality. Establishment of ideal minimum transmission power in each communication channel for the forward link and reverse link is the main objective of power control in a CDMA network. Another problem, known as the 'near–far problem', occurs due to the random distribution of the MS within the coverage area of a BS. Some MSs may be near to a BS and some may be far from a BS. For the same mean transmitted power, the BS may receive a much stronger signal from the nearer MS than that from the distant MS. So the main idea of power control schemes is to set MSs' transmission power to a minimum, so that BSs receive similar levels of signal from the MSs.

There are two types of power control techniques—**Open Loop Power Control (OLPC)** and **Closed Loop Power Control (CLPC).**

OLPC is usually implemented to give an estimate of minimum transmission power used for access channels during system access state, for example, RAChs (Reverse Access Channel), EAChs (Enhanced Access Channel) and RCCChs (Reverse Common Control Channels). The quantity E_c/I_0, the ratio of chip energy to interference plus noise spectral density is measured by the MSs per active candidate pilot set of the current carrier for the estimation of the reverse link transmission power.

CLPC is usually implemented for traffic channels that need a higher performance power control system. The MS initially uses the transmission power of the last message obtained from the access probe transmitted on the system access state. The BS determines a threshold (set point) power level to be ideal for receiving signals within a defined QoS (depending on FER). This set point is periodically adjusted. Through this power control mechanism, the BS instructs the MSs individually to increase or decrease transmission power according to the set point value. For CDMA IS-95 systems, power control method is applied to RACHs and FTChs and RTChs (forward and reverse traffic channels).

10.5 HANDOFF PROCESS IN A CDMA SYSTEM

Because of the user mobility and roaming across the networks, handoff is necessitated. Handoff from one BS to another may also require eliminating the shadow areas within the network service areas. Consider Fig. 10.7 where an MS is moving from one BS to another within a CDMA system network. When in the cell 1, the MS receives a signal from BS1, and acquisition and synchronisation are done to the Forward Pilot Channel (FPiCh) transmitted by BS1. But during the roaming process, the MS moves out to the coverage area of BS2 and needs to undergo a process for acquisition and synchronisation of FPiCh transmitted by BS2. This process is called *handoff process*. The handoff in a CDMA system is different from TDMA based GSM system.

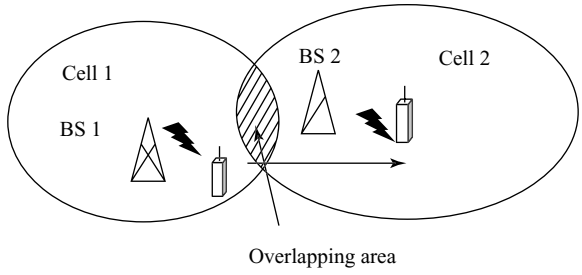


Fig. 10.7 *A handoff scenario—MS moves from BS1 to BS2*

10.5.1 Maintenance of Pilot Sets

To understand the handoff process in CDMA systems, understanding of the maintenance of a pilot signal is very much required. Each MS measures and internally stores information about the FPiChs present in the current location area during the hand-off process. This information is kept in four pilot sets that are regularly updated by the MS. The pilot signal may be the frequency and the forward channel sequence offset. There are four pilot sets in CDMA systems:

1. **Active pilot set:** During a call, this pilot set is associated with forward traffic channels assigned to the MS.
2. **Candidate set:** This set holds pilots not in the active set but having enough power signals to become active pilots.
3. **Neighbor set:** This set has pilots those are not active or candidate but may be considered as candidate for hand-off process.
4. **Remaining set:** These are pilots that do not fall under the above category but are in the area where the MS is located; or the pilots that do not have enough signal power strength to be considered as candidate for handoff.

All pilot sets are communicated to the mobile terminal by the BS via the neighbor list message except the first active set used by the MS to acquire a CDMA system. Every MS is equipped with devices that support digital signal processing of forward link signals. Out of four sets of devices, three are grouped, called *Fingers*, and are used to provide synchronization and demodulating forward link multipath signal components. The last device is called *Searcher*, used to search for pilots coming from other BSs in the area where the MS is presently located.

Both the BS and MS have deployed Rake receivers to receive diversity reception and can compare multipath signal components arriving from a single source. For different sources finger elements are used for the selection of the best traffic channel with the best quality. In a rake receiver, these fingers are used together to simultaneously demodulate the strongest multipath signal components received.

The CDMA handoff process is known as soft handoff. The MS uses a searcher to estimate the pilot's signal levels to evaluate E_c/N_0 (energy per chip/thermal noise plus the interference spectral densities). The searcher also evaluates the PN-offset (pilot sequence offset) of the received pilot. The MS always initiates the handoff process in the CDMA system that is generally called *mobile assisted handoff*. The six main parameters related to CDMA handoff are described below.

T_ADD This is used to control movement of pilots exceeding the E_c/N_0 target expressed in steps of 0.5 dB. The qualified pilot signals are moved from neighbor sets or from a remaining set to the active or candidate set.

T_DROP and T_TDROP These parameters are expressed in seconds that define the threshold detection value and drop timer for pilots on active and candidate sets respectively.

T_COMP Again expressed in units of 0.5 dB, it controls the transfer of pilots from the candidate to the active set. When the E_c/N_0 of a pilot signal in the candidate set exceeds an active pilot set by the value multiple of T_COMP then the candidate pilot replaces the weakest pilot of the active set.

NEIGH_MAX_AGE An AGE timer is initiated for the pilot moved to a neighbor set and incremented regularly up to NEIGH_MAX_AGE. It defines the maximum retention time for the pilot in a set, and if it expires, the pilot becomes the remaining set.

SRCH_WIN_A, SRCH_WIN_N and SRCH_WIN_R The pilot search window size is defined by these parameters expressed in chip intervals and are specified for each type.

ADD_INTERCEPT and DROP_INTERCEPT These are used to add or drop a pilot from the active set. These parameters are used to control transfer of pilots between the active and candidate sets.

SOFT_SLOPE This is used to determine pilots' E_c/N_0. It is defined as the slope in an inequality criterion to add or drop pilots from the active set.

10.5.2 Soft Handoff Process in IS-95 Networks

All the handoff parameters are received by the MS in an FPCh system parameters message, or during a call in the traffic system parameter message. When an MS receives handoff direction messages (General/ Extended/Universal) through the traffic channel, it immediately updates the pilot sets. Active participation of MSs in the handoff process is involved. The MS detects forward pilot channels (FPiCH) with an E_b/N_0 (energy per bit/noise plus interference spectral power density) greater than T_ADD and informs the MSC through a Pilot Strength Measurement Message (PSMM) or Extended PSMM. Based on this measurement information, the MSC identifies the target Base Station (BS) associated to these pilots and instructs the MS to perform handoff. The BS may also send the Neighbor List Update Message (NLUM) to the MS informing the preferential or natural neighbor pilots. After getting this NLUM, the MS must update the list set and set all the AGE timers corresponding to the pilots in the set. Fig. 10.8 illustrates the hand-off process and pilot set maintenance on CDMA IS-95 networks [1].

At the Position 1, P2 is the best pilot signal. The MS compares P1 with T_ADD during pilot set measurement and checks whether P1 exceeds T_ADD. The MS sends Pilot Strength Measurement Message (PSMM) for P2 to the server BS and the corresponding PN-offset requesting handoff and MS transfers pilot P1 to the candidate set.

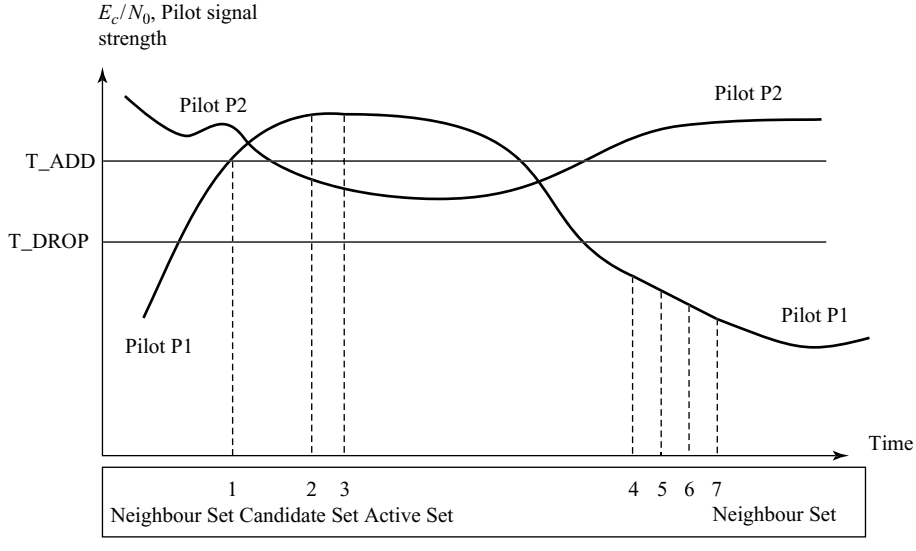


Fig. 10.8 *CDMA IS-95 hand-off process and pilot set maintenance*

At the point 2, P2 is the server pilot. Receiving the PSMM for P2, the MS informs associated MSC with the hand-off request; and a traffic channel is assigned to the MS. The MSC instructs the BS to send a handoff direction message to the MS for establishing handoff.

At the point 3, upon receiving the message from the BS, the MS transfers P1 to the active set and sends a Hand-off Completion Message during soft handoff process. The MS at this point will communicate with the two BSs using the traffic channel for both.

At the point 4, the MS finds that P1 drops below T_DROP and initialises a hand-off drop timer to this point. At the point 5, the T_DROP timer assigned to the pilot P1 expires and the MS again transmits PSMM to the BS. Upon receiving PSMM at the point 6, the MSC releases the allocated traffic channel informing the BS to transmit a Handoff Direction Message to the MS. Finally, at the point 7, the MS transfers P1 to the neighbor set, transmitting a Handoff Completion Message to the BS. The AGE timer is initialised at this point for P1.

It is to be mentioned here that for CDMA-to-CDMA soft hand-off process, the MS starts communicating with the target BS while connected to the old BS using the channels from both of the base stations. The MS can communicate more than two BS depending on the number of available fingers. Soft handoff generally occurs at the cell boundary regions when the received signal levels at MSs and BSs have poor quality because of the propagation effects. While the MS is in transition region, the call can be supported by signals through both cells, thereby eliminating the common border cell problems during ping-pong effect.

CDMA-to-CDMA hard handoff is also possible, where the MS disconnects the old BS and connects to the target. This happens when MS transitions from disjoint BS sets using different frequency assignments, or traffic channel frame offsets or band classes [1].

Summary

In this chapter the fundamentals of the CDMA technology has been described in respect of CDMA IS-95 networks. CDMA technology can increase the system capacity. Deployment of multiple correlators (RAKE receiver) in the CDMA systems enables tracking of individual signals coming from multipath propagation. The sum of the received signal strength is used to demodulate the signal. Again the simultaneous tracking of signals from different cells provides the underlying basis of soft handover in CDMA systems. Power control in CDMA systems is another important aspect to work effectively for the elimination of the 'near–far' problem. The RF power in the system needs to be controlled to avoid interference from the other mobile users within the cell and from neighbor cells. Understanding of the entire CDMA technology is quite complex and laborious. Readers are advised to read the references provided at the end of this chapter for further knowledge on CDMA technology.

References

[1] Leonhard, Korowajczuk *et. al*, *Designing CDMA2000 Systems*, John Wiley, 2004.

[2] Lee, Jhong Sam, and Leonard E. Miller, *CDMA Systems Engineering Handbook*, Artech House, Inc.,1998.

[3] CelTec/CelPlan, *CDMA IS-95 and cdma2000 Systems–Training Course*.

[4] Kumar, S., and S. Nanda, *High Data Rate Packet Communications for Cellular Networks using CDMA: Algorithms and Performance*, IEEE Journal on selected Areas in Communications, 17, pp. 472–492, 1999.

[5] Lee, W.C.Y., *Overview of Cellular CDMA*, IEEE Trans. on Vehicular Tech, Vol 40, May 1991.

[6] Kim, K.I., Ed, *Handbook of CDMA System Design, Engineering and Optimization*, Prentice Hall, USR, NJ 2000.

[7] C.S0005-B, *Upper Layer (Layer 3) Signaling Standard for cdma2000 Spread Spectrum Systems*, Release B, 3GPP2, April 2002.

[8] Salmasi, A., and K.S. Gilhousen, *On the System Design Aspects of Code Division Multiple Access Applied to Digital Cellular and Personal Communication Networks*, IEEE Vehicular Technology Confetence, pp. 57–62, 1991.

[9] Tiedemann, E.G., *CDMA2000-1X:New Capabilities for CDMA Networks*, IEEE Vehicular Technology Society Newsletter, Vol 48, Nov 2001.
[10] C.S0002-C, *Physical Layer Standard for CDMA2000 Spread Spectrum Systems*, Release C, 3GPP2, May 2002.
[11] Viterbi, A.J., *CDMA-Principle of Spread Spectrum Communication*, Addison-Wesley, Massachusetts, 1995.
[12] Wong, D., and T.J. Lim, *Soft Handoffs in CDMA mobile system*, IEEE Personal Communications, Vol. 4, pp. 6–17, June 1997.

Questions for Self-Test

10.1 Spread Spectrum transmission of data occupies a larger bandwidth than necessary.
a. True b. False

10.2 CDMA does not allocate frequency or time in user slots but gives the right to use both to all users simultaneously.
a. True b. False

10.3 Both of the PN sequences used at transmitter and receiver has to be identical and synchronized in spread spectrum technology.
a. True b. False

10.4 CDMA networks usually employ full-duplex access methods, such as Frequency Division Duplex (FDD).
a. True b. False

10.5 The reverse link is asynchronous due to the distinct signal propagation delay from the mobile stations to the base stations.
a. False b. True

10.6 The reverse channel structure allows up to
a. 62 different traffic channels and 32 different access channels.
b. 32 traffic channels, 7 pilot channels.

10.7 Does hard handoff occur in CDMA systems?
a. Yes b. No

10.8 IS-95 allows each user within a cell to use the same radio channel and users in adjacent cells also use the same radio channel.
a. True b. False

10.9 The CDMA system eliminates the need for frequency planning**.**
a. True b. False

10.10 The first power control is needed to mitigate the fading effect.
a. True b. False

10.11 Establishment of the ideal minimum transmission power in each communication channel is the objective of power control.
a. True b. False

10.12 The spreading process for forward and reverse links in IS-95 is different.
a. True b. False

10.13 IS-95 architecture provides base station diversity during soft handoff.
a. False b. True

10.14 Pilot channel is provided on the forward link to characterize signal condition.
a. False b. True

10.15 The RAKE receiver is used to resolve and combine multipath propagation effect.
a. True b. False

10.16 The PNLC is used in the _____channel for data scrambling.

10.17 To distinguish user-specific characteristics, the PNLC uses a unique code with _____.

10.18 Every CDMA channel in any BTS is identified by two parameters:_____ and _____.

10.19 A pilot channel is transmitted in a forward channel at a _____ power than the user's channel.

10.20 Forward traffic channels are used for _____.

10.21 _____ WALSH Code used for SynCh.

10.22 The reverse link is _____ due to the distinct propagation delay.

10.23 In IS-95, data for FCh is grouped into _____ms.

10.24 The PN offset is used for both forward and reverse links.
a. False b. True

10.25 PN offset can be used in distance mapping from MS to BS.
a. True b. False

10.26 In CDMA, handoff is always mobile assisted.
a. False b. True

10.27 In CDMA, each mobile in a given cell is assigned a different spreading code. Why?

10.28 What are the advantages and disadvantages of the CDMA system?

10.29 How many types of mask, are there in CDMA system? What are their purposes?

10.30 Discuss about the forward and reverse links in IS-95 systems.

10.31 What are the different PN sequences used in the CDMA IS-95 networks? Discuss the importance of those PN sequences.

10.32 How are different channels in the forward link identified in an IS-95 system?

10.33 Why is power control so important in CDMA systems? Discuss about the two methods of power control mechanisms in IS-95 systems.

10.34 What are the different stages of pilot signal maintenance? How is it very important in handling handoff mechanism?

10.35 With an example, discuss, the soft handoff mechanism in CDMA IS-95 networks.

10.36 Point out the evolutionary path of CDMA system development.

10.37 In the DS-CDMA system, the receiver needs to accurately synchronize the locally generated PN sequence at the receiver with the incoming PN sequence waveform—discuss the reason.

10.38 What is called near-far problem? How is this problem solved in a fading environment?

10.39 The IS-95 system uses a rate 1/2 convolutional encoding in forward link and a rate 1/3 convolutional coding in reverse link. Why?

10.40 How is a call processed in IS-95 networks?

10.41 In a CDMA cellular system, each mobile station monitors pilot signals from the serving base station and other neighbor base stations. MS uses a searcher to estimate the pilot's signal levels to evaluate E_c/N_0. Describe the six parameters related to the CDMA handoff process.

3G–The Universal Mobile Telecommunication System (UMTS)

11

Introduction

The aim of future networks is to support voice, video and data together. Technology is now geared in the direction towards the development of third generation (3G) networks to support multimedia communications. The driving technology obviously will be based on Internet Protocol (IP). The boom of mobile services supposed to be available in future will generate revenue from the ever-increasing number of mobile subscribers. The success of GPRS networks in providing limited data services is the key driving force needed to deploy the IP-based core network for 3G, fitting into the main components of the GPRS network and the existing Global System for Mobile Communication (GSM) infrastructure.

As described in Chapter 9, for a GPRS network, two new nodes, namely SGSN and GGSN, are introduced to support packet data with the external IP networks. This brings the IP protocol as the transport mechanism between SGSN and GGSN allowing email and web-based services.

The Universal Mobile Telecommunication System (UMTS) is a third-generation (3G) wireless system that delivers high-bandwidth data and voice services to mobile users. UMTS evolved from GSM and has a new air interface based on Wideband Code Division Multiple Access (WCDMA).

Release 99 (R99) is the first version of UMTS architecture based on the new multiple access technology WCDMA for increased utilisation of radio resources. A part of the GSM/GPRS network can still be used in addition to the new components. The Third Generation Partnership Project (3GPP) has specified the R99 standards.

The idea of Release 4 (R4) is to merge the circuit-switched network and packet-switched network into a single entity. The main change in the network is the separation of the tasks in the circuit-switched core network.

UMTS Release 5 (R5) has a control layer that is responsible for handling the signaling for multimedia sessions. The new system is called IP Multimedia Subsystem (IMS). The IMS is an extension of the packet-switched network.

The evolutionary path for 3G UMTS networks with their architectural details and functionality will be provided in this chapter. The concept of an all-IP network evolving from UMTS releases will also be given along with a summary of the impact of mobile data services of this network.

11.1 UMTS NETWORK ARCHITECTURE–RELEASE 99

The first deployment of the UMTS R99 network architecture is shown in Fig. 11.1. The major change is in the Radio Access Network (RAN) based on WCDMA and Asynchronous Mode of Transmission (ATM).

The UMTS architecture defines three main functional entities:

User Equipment (UE), UMTS Radio Access Network (UTRAN) and the Core Network (CN), which are again divided into circuit-switched (CS) and packet-switched networks (PS). A part of the GSM/GPRS network can still be used with some additional new components to be implemented for UMTS.

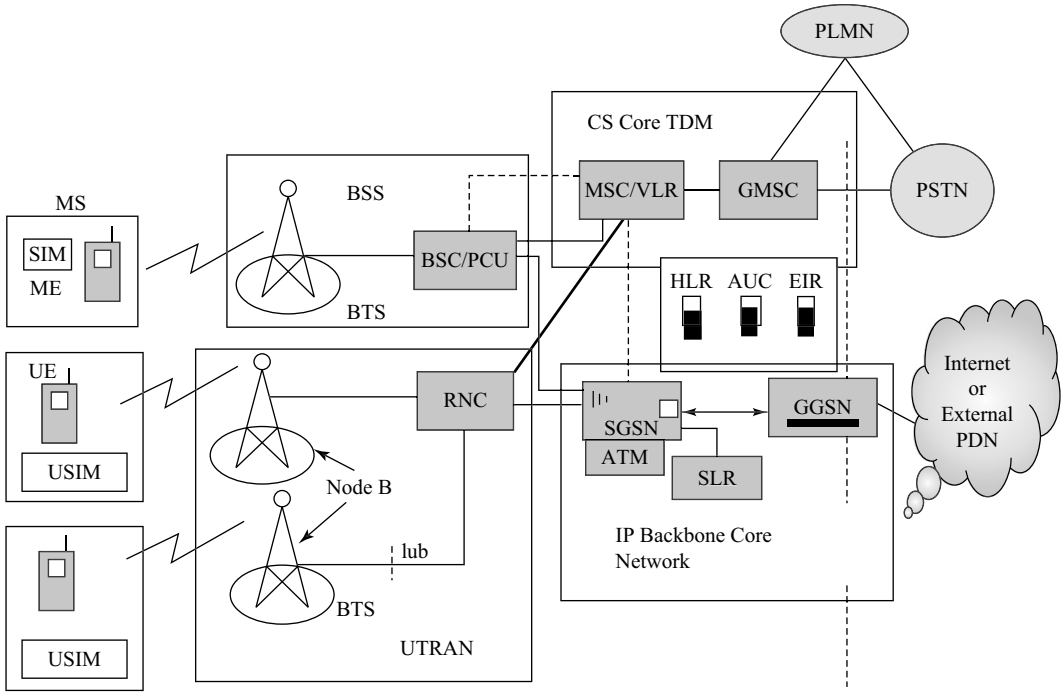


Fig. 11.1 *UMTS networks architecture*

User Equipment (UE) UE replaces the MS for GSM/GPRS networks. A subscriber must buy a new handset for 3G services with a new SIM called USIM. Every UE may contain one or more USIM simultaneously. USIM is a user subscription to the UMTS mobile network and contains all relevant data that enables access onto the subscripted network. The main difference between a USIM and a GSM SIM is that by default, a USIM is downloadable and can be accessed via the air interface and be modified by the network. The USIM is a universal integrated service card having much more capacity than the GSM SIM. It can also store JAVA applications. The Mobile Terminal (MT) for UMTS may be single-radio mode or multi-radio mode.

Node B The base station used in UMTS is known as 'Node B' that replaces BTS. It provides the physical radio link between the UE and the network. As the access technology is different from GSM/GPRS, Node B is capable to handle CDMA subscriber on the new frequency bands. It can also support higher data rates used for 3G. Node B is the termination point between the air interface and the transmission network of the RAN. It performs the necessary signal processing functionalities for the WCDMA air interface and is more complex than BTS. Node B is responsible for the following:

1. **Power control:** It measures the actual signal-to-interference ratio (SIR), compares it with the threshold value and then may trigger the change of transmitting power of UE.
2. **Reports the RNC (Radio Network Controller):** The measured values are reported to RNC.
3. **Combines the received signals coming from multiple sectors of the antenna that a UE is connected to:** It converts the signals into a single data stream before it transmits to the RNC. This may help to soften the handover procedure for UMTS networks.

Three types of Node B are possible—UTRA-FDD Node B, UTRA-TDD Node B and Dual Mode Node B, supporting both FDD and TDD modes.

Radio Network Controller (RNC) The RNC is the main element in the Radio Network System (RNS) and controls the usage and reliability of radio resources. An RNC is similar to a BSC and is interfaced with the CS of a GSM core network (MSC) in order to handle circuit-switched calls along with SGSN for packet data transport. It also needs to be capable of supporting interconnections to other RNCs, which is a new feature of UMTS. All the decision-making processes are done here and is software based. RNC decides, as does the PCU in a GPRS, to route the call to the MSC for voice call or to SGSN for data packets. The main tasks for RNC are call admission control, radio bearer management, power control and general management controls in connection to OMC (Operation Management Control).

There are three types of RNCs—Serving RNC (SRNC), Drift RNC (DRNC) and Controlling RNC (CRNC). The SRNC controls a user's mobility within a UTRAN. It is a connection point to the core network towards MSC or SGSN. The DRNC receives connected UEs that are drifted or handed over from the SRNC cell connected to a different RNS. The RRC (Radio Resource Controller) is still connected to SRNC. The DRNC then exchanges the routing information between the SRNC and UE. Thus, the DRNC provides radio resources to the SRNC to allow soft handover. CRNC controls, configures and manages an RNS and communicates with the Node B application part (NABP) protocol with the physical resources of all Node Bs connected via the Iub interfaces. Any access request from the UE is forwarded to the CRNC.

It is to be noted that a 3G MSC is not exactly similar to a 2G MSC, which is a narrowband device and is connected to the BSS via an 'A' interface. The traffic data rate is 64 kbps (for voice—64 kbps Pulse Code Modulation, PCM). For 3G RAN, it has to handle speech across the ATM over the interface for circuit-switched connection with adaptive multirate. So, a new node called Inter-working Function (IWF) is needed in between the RAN and MSC. IWF is responsible for transcoding speech into 64 kbps PCM and vice versa. Secondly, for control signaling, the IWF helps in mapping the MSC message signaling into RAN message signaling.

11.2 UMTS INTERFACES

Apart from the interfaces used in GSM/GPRS networks, a number of other new interfaces are defined for the implementation of UMTS services based on the new nodes for UMTS, namely RNC and Node B. The reason for defining interfaces is to operate the different UEs from different network operators accessing the UTRAN. The new interfaces that are introduced are as follows.

Uu The radio interface between UE and Node B.

Iub Interface between Node B and RNC.

Iur Interface between RNC and RNC.

Iu (CS) Interface between the RNC and MSC for circuit switching.

Iu (PS) Interface between the RNC and the SGSN for packet switching.

Figure 11.2 shows the different interfaces used in a UMTS network. All the new radio interfaces for UMTS are characterized through protocols that are divided into two groups; *user plane protocols* to carry user data and *control plane protocols* for controlling the connection between the UE and the network nodes.

All these interfaces are standardised on ATM for Layer 2 except the Uu interface. ATM offers good QoS for all applications. Interface Iub transports both circuit and packet-switched data. The Iu (PS) transport packet-switched data is based on IP with the ATM adaptation layer 5. But the interfaces Iur and Iu (CS) transport circuit-switched connection based on ATM adaptation layer 2, suitable for real-time applications.

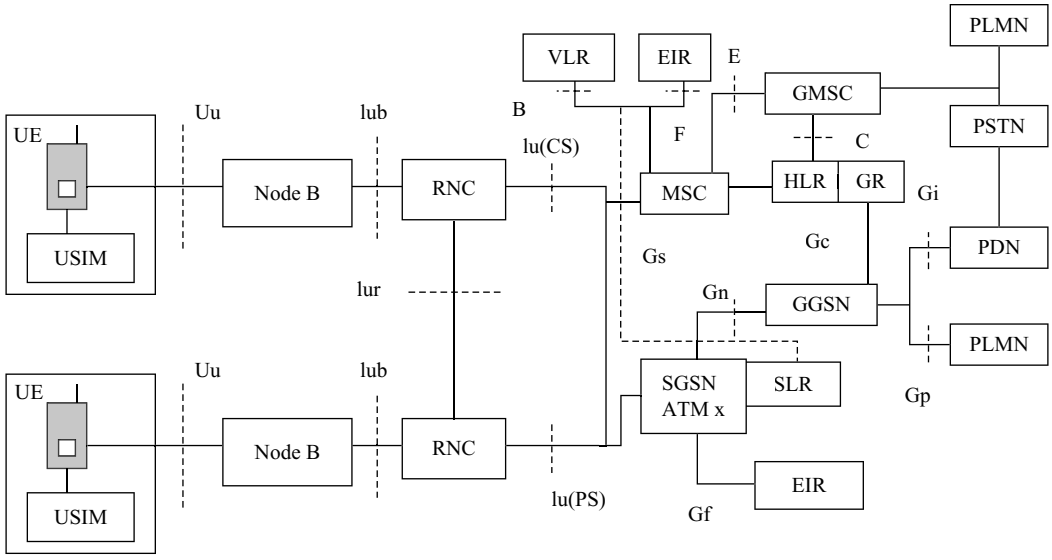


Fig. 11.2 *UMTS interfaces*

11.3 UMTS NETWORK EVOLUTION

The next evolutionary steps for UMTS architecture from R99, is the Release 4 (R4) that merges the separate circuit (MSC/VLR, GMSC) and packet switched (SGSN/SLR, GGSN) data into one based on IP network infrastructure, and also supports voice-over IP (VoIP) technology. The concept is not to modify the PS domain and radio interface and service set, but to enable operators to use IP or ATM as a common means to reduce cost. The task of transporting real-time applications over IP with good quality of service is more complex than ATM. R4 proposes a solution to support VoIP over UMTS. To apply this solution, new nodes like Mobile-Services Switching Center (MSC-Server), Media Gateway (MGW) and Gateway Mobile-Services Switching Center (GMSC) servers are required.

MGW is responsible for user traffic handling inside the core network. It also converts the protocols for radio subsystem for the fixed network PSTN or pre Release 4 PLMN. MGW contains some stream-manipulating functions for the adaptation of CS voice traffic to VoIP.

The GMSC and MSC server provide call control and mobility management functionalities of the GSM/UMTS network. Figure 11.3 shows the UMTS R4 architecture. The idea is to use ATM or IP in between MGW. For IP, SGSN and MGW can be reduced to a single element responsible for carrying IP traffic, both for real and non-real time application. The core network for R4 consists of two parts—IP-based traffic network and SS7 signaling network. The GMSC server and MSC server are connected to MGW through H.248 protocol, and the MGWs are connected through RTP/UDP/IP protocols. Real time protocol (RTP) transports real time applications over a packet-switched network. User Datagram Protocol (UDP) is used for unreliable data transmission.

Two additional gateways are required within the IP-based core network in Release 4. These are:

1. Transporting Signaling Gateway (T-SGW) is used to convert call-related signaling such as call set-up and call release between the PSTN (SS7) or pre Release 4 PLMN (SS7) and Release 4 network
2. Routing Signaling Gateway (R-SGW) performs signaling conversion for roaming, mobility management between the SS7 based signaling of pre Release 4 network and IP-based signaling of the Release 4 network

Some of the new interfaces are needed for these changes. Fig. 11.4 shows the new interfaces inside the Release 4 UMTS network with the additional nodes.

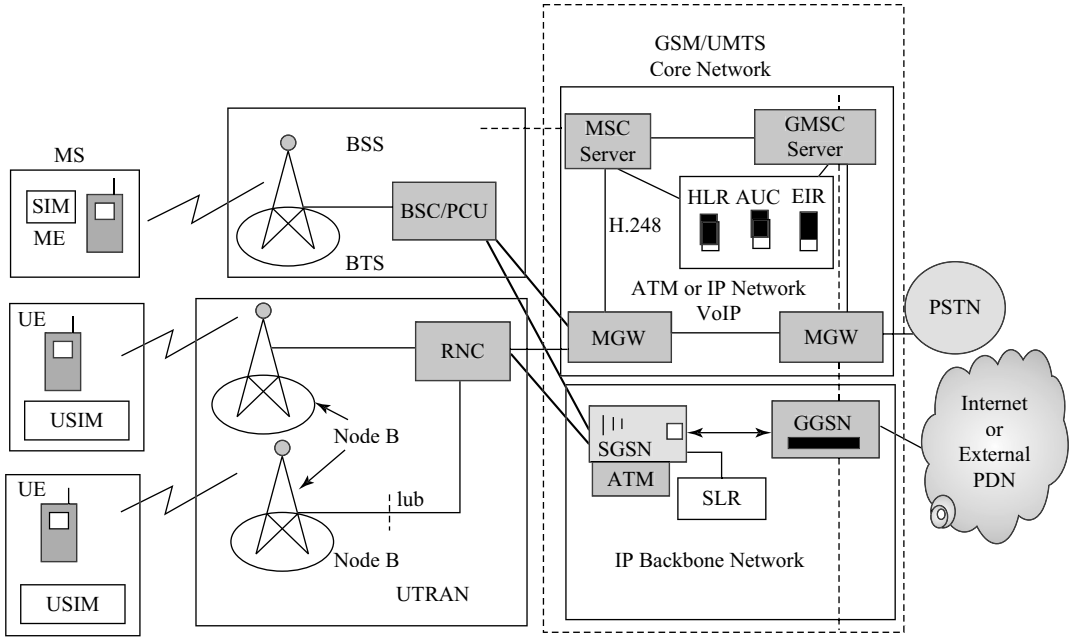


Fig. 11.3 *UMTS Release-4 architecture*

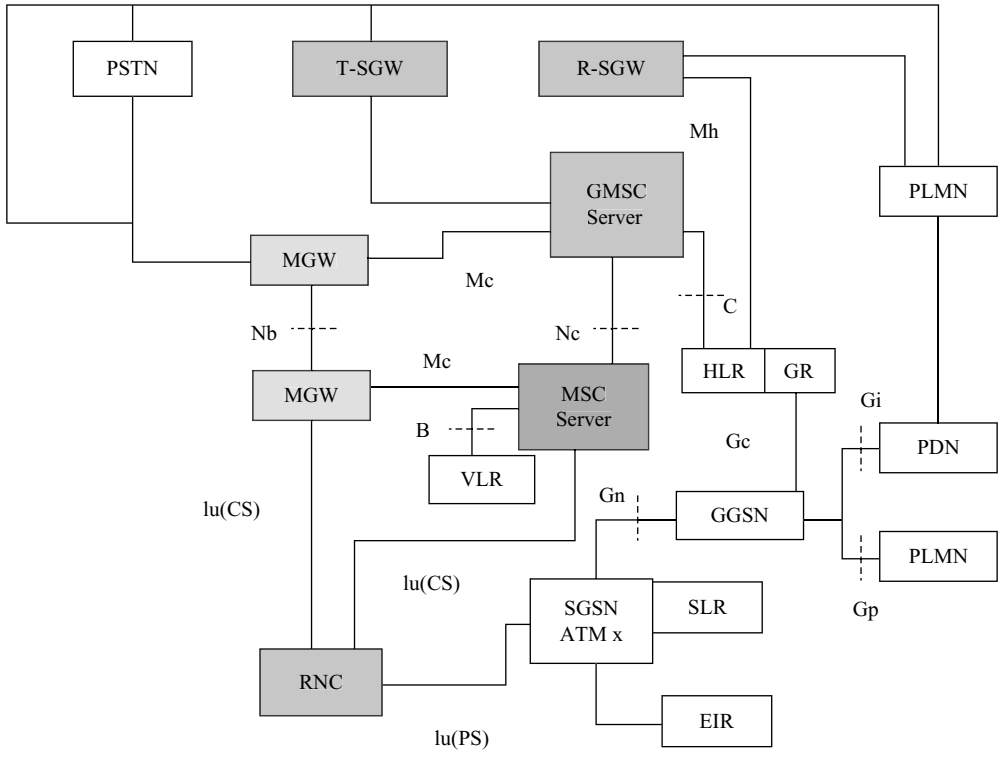


Fig. 11.4 *UMTS new interfaces*

The interface Mc works between the MGW and MSC servers. The interface Nc is either a classical SS7 interface or is IP based. For IP-based services, several other protocols need to be implemented.

The BICC (Bearer Independent Call Control) or SIP (Session Initiation Protocol) work between the MSC server and the GMSC server. The BICC supports narrowband ISDN service over broadband backbone network, ATM or IP.

Stream Control Transmission Protocol (SCTP) is necessary for transporting SS7 message on an IP interface. On the Mh interface, SS7 is implemented over SCTP/IP.

11.4 UMTS RELEASE 5

The primary aim of Release 5 is to provide IP-oriented services by the operators, unlike the core network modification of Release 4 for IP transport. It is expected to generate revenue by providing IP multimedia services. To meet the demand, conventional circuit-switched connections in the CS domain have to be replaced by enhanced IP-based access completed by packet switching. This is realized by the IP Multimedia Subsystem (IMS) that overlays the existing architecture. The IMS is an extention of packet-switched core network intended to offer access independence and inter-operation with wireline terminal access across the Internet. The interfaces therefore specified to conform as far as possible to IETF 'Internet Standards' where an IETF protocol is to be used. Fig. 11.5 shows the UMTS Release 5 network architecture using IMS.

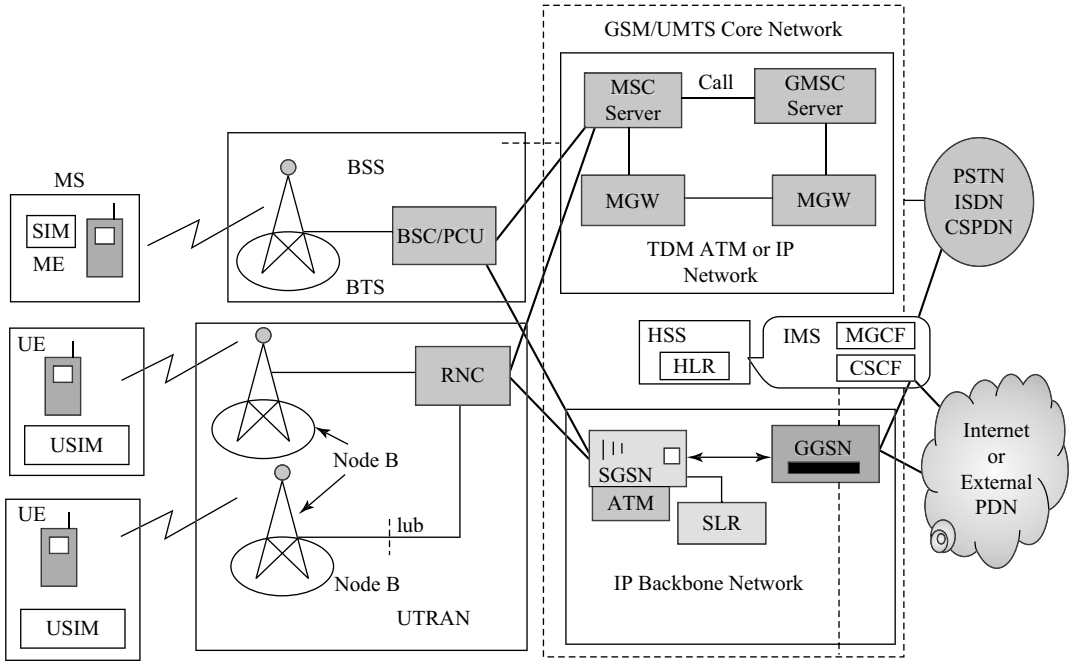


Fig. 11.5 *UMTS Release 5 networks*

The IMS uses PS domain to transport multimedia signaling and bearer traffic. For the future VoIP services, it is a prerequisite that all data of a multimedia service pass through the same core network. There are two components of IMS—one, Call State Control Function (CSCF) which is an entry point for signaling of incoming calls; and two, Media Gateway Control Function (MGCF), responsible for inter-working with the PSTN of CS domain. UMTS Release 5 introduces an integrated database called Home Subscriber Server (HSS), which provides the subscriber profile information. CSCF interacts with HSS for database queries for location management and routes the call accordingly.

The Session Initiation Protocol (SIP) has been standardised as the call and session control protocol for the IMS. SIP is a signaling protocol defined by IETF that establishes sessions, modifies and releases

the sessions over IP networks. SIP does not transport session data but only the information regarding the session, name, time elapsed or transport protocol. The simplified UMTS architecture is given in Fig. 11.6.

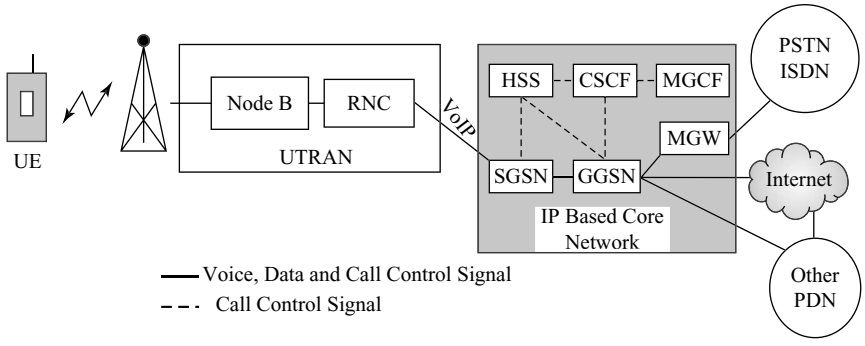


Fig. 11.6 *Simplified UMTS network*

Clearly, the transport layer nodes perform voice and data transport.

Node B Base transceiver station (BTS)
RNC Radio network controller or Base station controller (BSC)
SGSN Serving GPRS support node
GGSN Gateway GPRS support node
MGW Media gateway

The call control layer nodes mainly perform the call control function.

CSCF Call State Control Function
MGCF Media Gateway Control Function
HSS Home Subscriber Server

11.5 UMTS FDD AND TDD

The physical layer (L1) access for UMTS is based on Wideband Direct-Sequence Code Division Multiple Access (WCDMA) technology with two duplex modes:

FDD (Frequency Division Duplex) and TDD (Time Division Duplex).

A physical channel in FDD is defined by the code and frequency, whereas in TDD it is defined by code and time slot.

FDD Uplink and downlink transmissions use separate frequency bands like GSM by placing different frequency channels. ITU-T spectrum usage for FDD is 1920–1980 MHz for uplink traffic and 2110–2170 MHz for downlink traffic as shown in Fig. 11.7. The minimum paired frequency allocation required is 5 MHz, one for uplink and another for downlink. The frequency separation between uplink and downlink is 190 MHz. However, for better coverage and services, an operator needs 3–4 channels (2×15 MHz or 2×20 MHz) to be able to build a high-speed, high-capacity network.

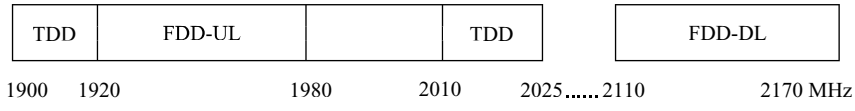


Fig. 11.7 *UMTS frequency allocation*

TDD TDD system, on the other hand, requires only one 5 MHz unpaired band to operate. Time division multiplexing, i.e., allocating different time slots, separates the uplink and downlink traffic in TDD. The transmitter and receiver are not separated in frequency.

Satellite uplink and downlink is 1980–2010 and 2170–2200 MHz.

A UTRA Absolute Radio Frequency Channel Number (UARFCN) designates carrier frequencies. The general formula, relating frequency to UARFN, is

UARFCN = 5 * (frequency in MHz).

The difference between FDD and TDD is only on the lower layer radio interface. On the higher layers, the two systems are the same. The 10-ms frame structure is divided into 15 equal time slots of 2560 x T_c (T_c = chip rate) to be allocated either in uplink or downlink. With such flexibility, the TDD scheme becomes more adaptive between varied environment and deployment scenarios. In many cases, at least one time slot is allocated for downlink and at least one time slot for uplink.

11.6 UMTS CHANNELS

The UMTS channels at the UE are mainly divided into three hierarchical layers. These are logical, transport and physical channels (Fig. 11.8). The logical channel is a particular type of information over the radio interface that is mapped into the transport layer using MAC protocol. The classification of transport channel depends on how the information is transported between the UE and RNC. The physical channel is the actual channel on the air interface based on WCDMA code and frequency. Each physical channel is identified by its frequency, spreading code, scrambling code and phase of the signal.

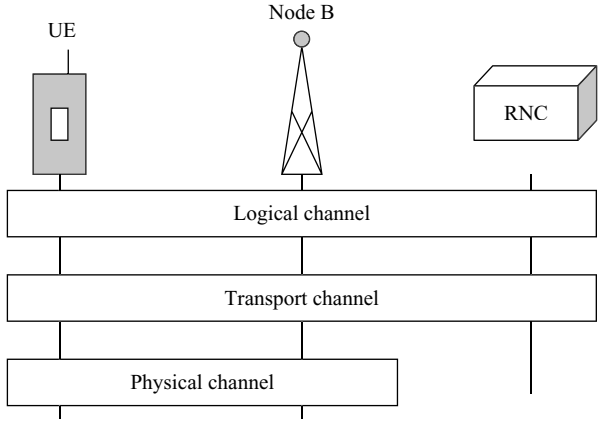


Fig. 11.8 *UMTS three channels*

11.6.1 Logical Channels

The MAC layer provides data transfer services on logical channels. Each logical channel type is defined by what type of information it transfers. A logical channel is again divided into control and traffic channels.

Control Channel Used to transfer control-plane information only.

BCC (Broadcast Control Channel) It is the downlink channel used to broadcast general information for the UE.

PCC (Paging Control Channel) Also a downlink channel, it carries paging information to warn the UE that there is a communication request.

CCC (Common Control Channel) This is both a downlink and uplink channel, and carries information from the network to UE without any dedicated channel.

DTCH (Dedicated Traffic Channel) Bi-directional and point-to-point channel, dedicated to one UE, for the transfer of user information.

Traffic Channel Used for the transfer of user plane information only.

CTCH (Common Traffic Channel) Carries dedicated user information among the selected UEs in the downlink direction.

DCCH (Dedicated Control Channel) Bi-directional, point-to-point channel that transmits dedicated control informations to the network and a UE.

11.6.2 UMTS Downlink Transport and Physical Channels

Transport channels are used for information transfer from lower layers to higher layers. A transport channel is defined by the manner and specific characteristics by which it transfers data over the air interface.

BCH (Broadcast Channel) Transports BCCH. It is then carried through the primary common control physical channel (P-CCPCH).

PCH (Paging Channel) Transports the PCCH and is carried in the secondary common control physical channel (SCCPCH).

FACH (Forward Access Channel) Transports a number of logical channels carrying common and dedicated control and traffic, such as CCCH, CTCH and DCCH, DTCH.

DCH (Dedicated Channel and DSCH) Downlink shared channel, dedicated control and traffic can be transported on these channels; statistical multiplexing is done to the multiple uses on one of the transport channel.

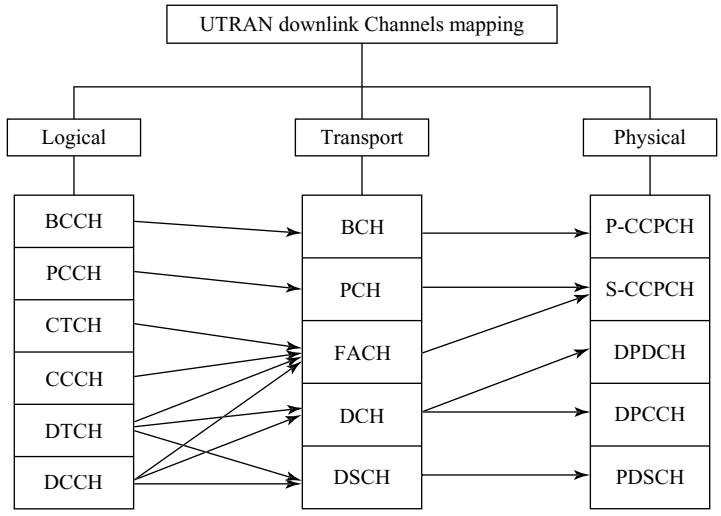


Fig. 11.9 *UTRAN downlink channel mapping*

11.6.3 UMTS Uplink Transport and Physical Channels

Figure 11.10 shows the chart for UTRAN uplink transport and physical channel mapping.

CPCH (Common Packet Channel) is used for data traffic which is carried through the Physical Common Packet Channel (PCPCH).

In the uplink, the Random Access Channel (RACH) is present for the transport of dedicated traffic as well as common and dedicated control information. Fig. 11.11 is the representation of random access channel.

RACH is an **uplink** transport channel and is always received from the entire cell. The RACH is characterised by a collision risk and by being transmitted using open loop power control.

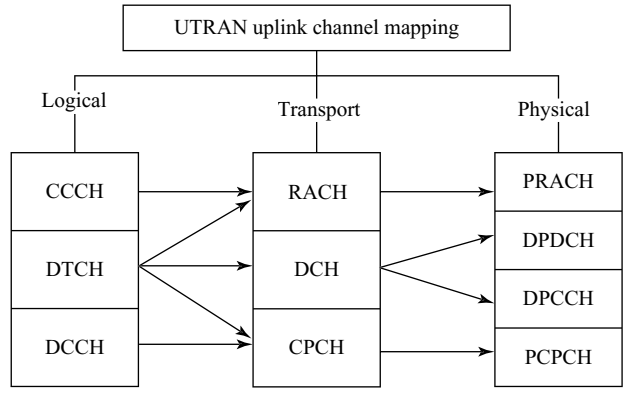


Fig. 11.10 *UTRAN uplink channel mapping*

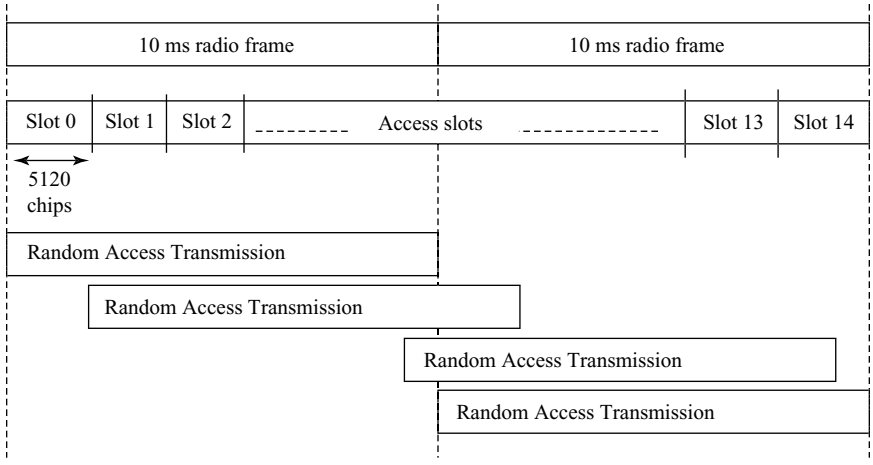


Fig. 11.11 *UMTS random access methods*

Figure 11.12 shows the message structure of an RACH. The 10-ms RACH message part radio frame is split into 15 slots, each of length $T_{slot} = 2560$ chips. Again, each slot consists of two parts, a data part to which the RACH transport channel is mapped and a control part that carries Layer 1 control information. The data and control parts are transmitted in parallel.

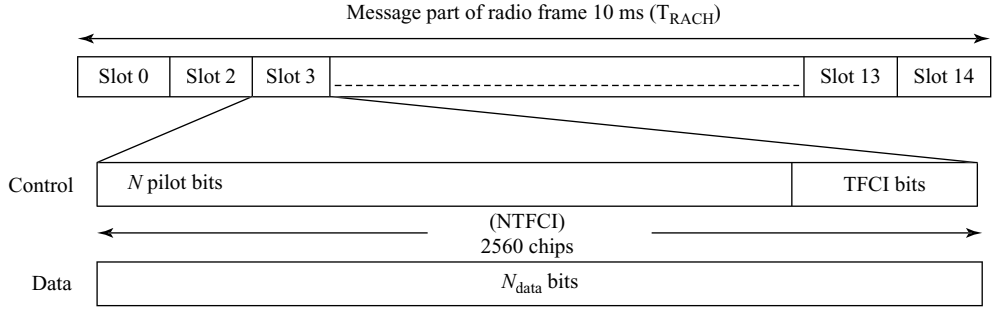


Fig. 11.12 *RACH Message structure*

A 10-ms message part consists of one message-part radio frame, while a 20-ms message part consists of two consecutive 10-ms message part radio frames. The data part consists of $10*2^k$-bits, where $k = 0, 1, 2, 3$. This corresponds to a spreading factor of 256, 128, 64, and 32 respectively for the message data part. Transport Format Combination Indicator (TFCI) contains the information relating to data rates. Pilot bits are always the same and are used for channel synchronisation.

11.7 UMTS TIME SLOTS

UMTS has several different time-slot configurations depending on the used channel. Figure 11.13 explains the downlink time-slot allocation for DPDCH (Dedicated Physical Channel) and Fig. 11.14 explains the uplink time-slot allocation.

Dedicated Channels (DCH) are transported at the physical layer using the dedicated physical channel (DPDCH). The physical layer channel has a control channel associated with it, the DPCCH (Dedicated

Physical Control Channel), which contains relevant information with regard to the physical transmission of data. DPDCH and DPCCH are separated using different channelization codes for uplink transmission.

TPC stands for Transmit Power Control; Feedback Information (FBI) is used for closed loop transmission diversity. TPC is used to control the power in the downlink. It is 1-bit information, 0 indicates that power needs to be decreased, and 1 indicates that power is up. FBI is used to apply diversity techniques in the BTS, getting feedback from the mobile device.

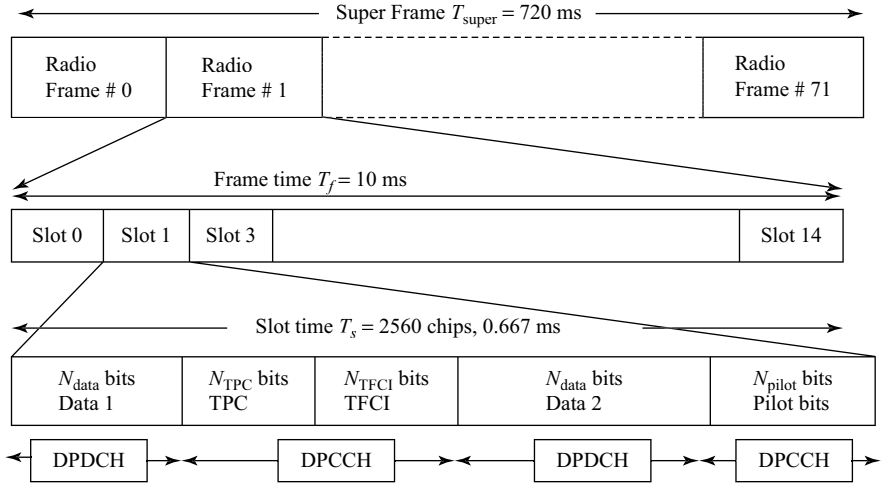


Fig. 11.13 *Downlink DPCH slot allocation*

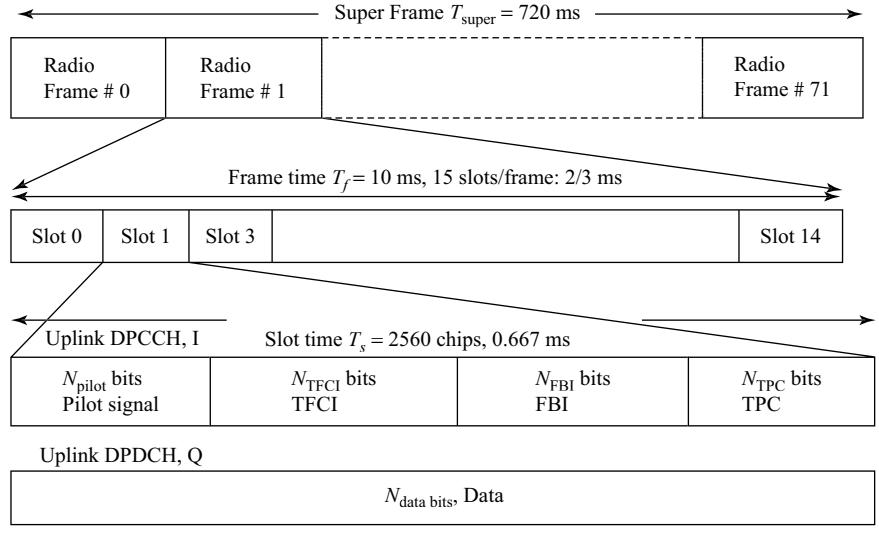


Fig. 11.14 *Uplink DPCH slot allocation*

DPDCH and DPCCH are combined in the downlink channels using TDM and transmitted under the same coded channel. This combined channel is known as the dedicated physical channel for downlink (DPCH). In the uplink, DPDCH and DPCCH are I/Q multiplexed together to avoid uplink interference.

11.8 UMTS NETWORK PROTOCOL ARCHITECTURE

The network protocol architecture for UMTS radio access network (UTRAN) is mainly divided into three layers. Each of the layers is again divided into control planes and user planes. Figure 11.15 shows the network protocol architecture for UMTS.

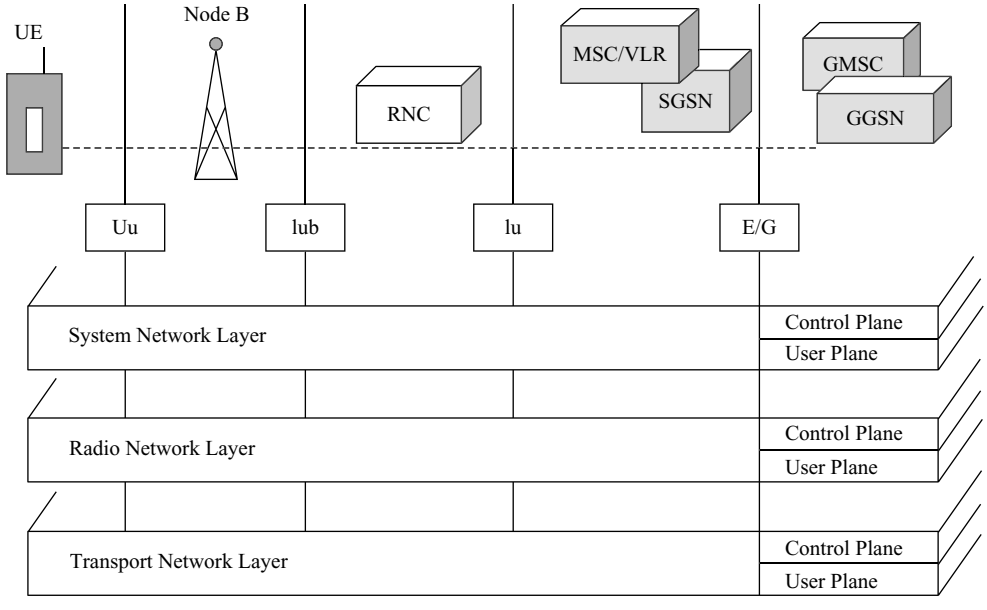


Fig. 11.15 *UMTS Network protocol architecture*

Transport Network Layer It allows communication between UTRAN and Core Network (CN).

Radio Network Layer Protocols and functions provide management of radio interface and communication between UTRAN components and between UTRAN and UE.

System Network Layer It allows communication between CN and UE.

The general protocol model architecture for UMTS is shown in Fig. 11.16. The structure is based on the principle that the layers and planes are logically independent of each other. Therefore, as and when required, the standardization body can easily alter protocol stacks and planes to fit future requirements.

The Control Plane includes the Application Protocols, i.e., RANAP (Radio Access Network Application Part), RNSAP (Radio Network Subsystem Application Part or NBAP (Node B application part), and the Signaling Bearer for transporting the Application Protocol messages.

Among other things, the Application Protocol is used for setting up bearers (i.e., Radio Access Bearer or Radio Link) in the Radio Network Layer. The User Plane includes the data stream(s) and the data bearer(s) for the data stream(s). The data stream is characterised by one or more frame protocols specified for that interface.

The **Transport Network Control Plane** does not include any radio network layer information, and is completely in the transport layer. It includes the ALCAP (Access Link Control Application Protocol) protocols that are needed to set up the transport bearers, i.e., data bearer for the user plane. It also includes the appropriate signaling bearer(s) needed for the ALCAP protocols. The transport network control plane is a plane that acts between the control plane and the user plane. The introduction of the transport network control plane is performed in a way that the application protocol in the radio network control plane is kept completely independent of the technology selected for data bearer in the user plane. Indeed, the decision to actually use an ALCAP protocol is completely kept within the transport network layer.

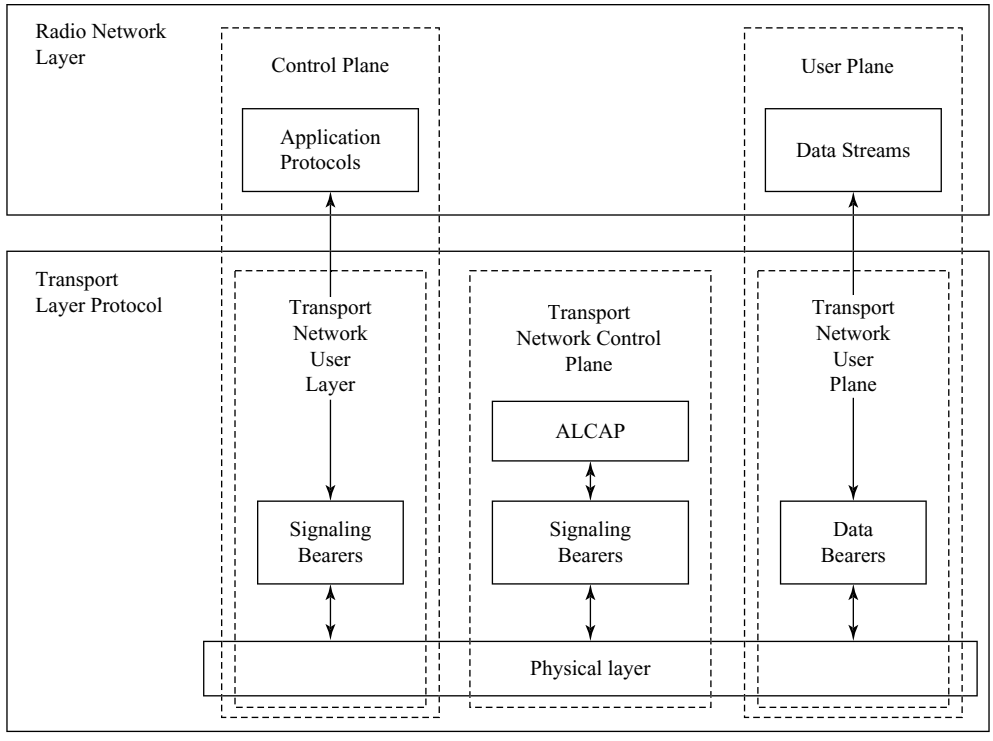


Fig. 11.16 *General protocol model*

11.9 UMTS BEARER MODEL

The procedure for a mobile device connecting to a UMTS network can be split into two areas—NAS (non- access stratum) and AS (access stratum)—as shown in Fig. 11.17. AS is related to the layers and subsystems that offer services to NAS. AS interacts with NAS through Service Access Points (SAP). AS basically consists of all the elements in RAN and underlying ATM networks, whereas NAS functions are between the mobile device and the core network, and the UTRAN provides seperation of AS and NAS functions. The major interfaces for UTRAN are Uu and Iu. Uu works between UE and UTRAN, and Iu works between UTRAN and core network. These interfaces are divided into user and control planes.

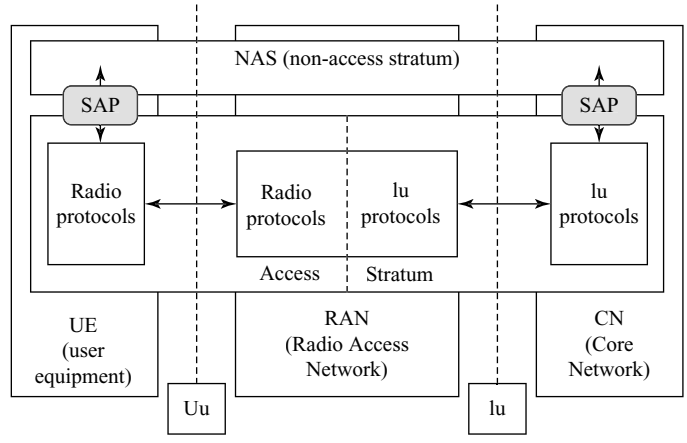


Fig. 11.17 *UMTS bearer model*

Radio Access Bearer (RAB) is the bearer service for a UMTS network. A bearer service is a means to establish a link between two points with specified characteristics. An RAB is defined as the service that the AS provides to the NAS for transfer of data between the UE and core network. An RAB consists of a number of subflows that are data streams to the core network, having different qualities of service characteristics.

RAB subflows are established and released at the time of RAB establishment and are released and delivered together at the transport bearer.

11.9.1 UMTS Interfaces

The UTRAN is subdivided into RNS (Radio Network Subsystems), each consisting of RNC and Node B. The interface between the RNSs is called the Iur interface which plays a key role in the handover process. All the Iu, Iub and Iur interfaces use ATM as the transport layer. So, RNC is seen as an ATM switch, whereas Node B is seen as the host accessing the ATM network. The Iub interface is a user-to-network interface (UNI) whereas the Iu and Iur interfaces are considered as the network-to-network interfaces (NNI). For each user connection to the core network, there is only one RNC that connects the UE to the core-network domain. This RNC is known as the serving RNC (SRNC). The SRNC and the Node B under this RNC together is known as SRNS. This is a logical definition used for that UE only. The controlling RNC within an RNS is called CRNC (Controlling RNC). This is with reference to the Node B, cells under its control and all the common and shared channels within it. Figure 11.18 shows the different interfaces between nodes.

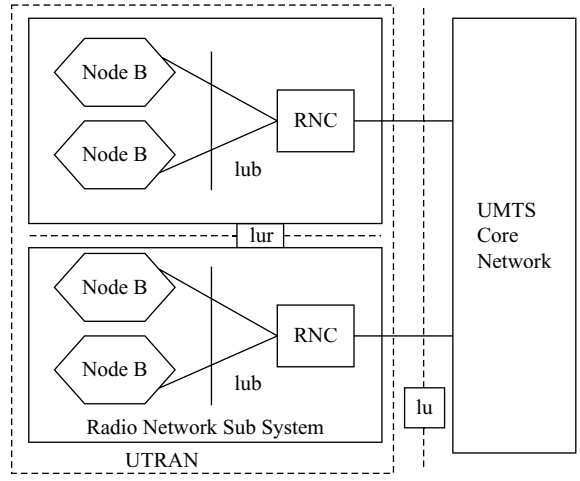


Fig. 11.18 *UTRAN Interfaces*

11.10 UTRAN TRANSPORT NETWORK

The UTRAN transport network consists of nodes and links for the purpose of transporting user, signaling and management traffic, supporting at the same time different levels of QoS. The solution for radio network control plane at the Iub and Iur interfaces is shown in Fig. 11.19.

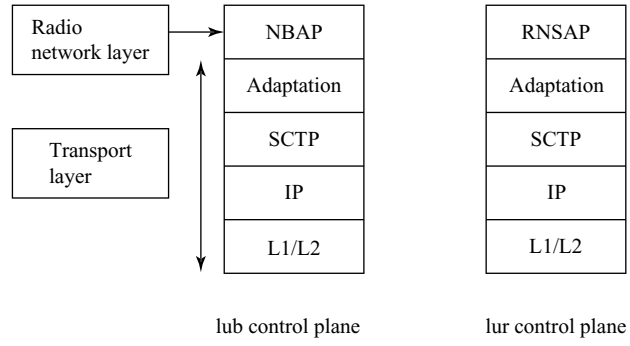


Fig. 11.19 *UTRAN control plane protocols*

The physical layer (L1) consists of WCDMA and ATM. The data-link layer (L2) is comprised of a number of sublayers, i.e., Media Access Control (MAC), Radio Link Control (RLC), Packet Data Convergence (PDCP) and the Broadcast/Multicast Control (BMC). The PDCP layer is only used for packet data such as IP between the UE and RNC, which is used for upper layer header compression.

For the Release 5 UMTS network, currently there are two alternative methods of radio network signaling a bearer for Iur transport. These are SCCP (Signaling Connection Control Part) or MTP3 user adaptation protocol, and SS7 SCCP user adaptation Layer SUA. The most preferable is M3UA for the radio network layer signaling bearer in the Iur interface.

SCTP is the IETF protocol designed to transport PSTN signaling messages over IP networks. It performs the following functions:

1. Reliable data delivery to the user; also notifies the loss of data packets,
2. Maintains in-sequence delivery of packets,

3. Bundling multiple user messages into a single SCTP packet.
4. Avoids head-of-line blocking offered by TCP and helps in message oriented data transfer.

M3UA is a protocol developed by IETF for the transport of any SS7 MTP3 user signaling such as ISUP, SCCP and TUP over IP using SCTP. It performs the following functionalities:

1. Support for transfer of SS7 MTP3 user part messages.
2. Support for the management of SCTP transport protocol between a signaling gateway and one or more IP-based signaling nodes to ensure transport availability to MTP3 user signaling applications.
3. Support for the seamless operation of MTP3 user protocol peers.
4. Support for the distributed IP based signaling nodes.
5. Support for the asyncronous reporting of status changes to management.

For the user plane, several solutions are provided which are based on the transport of IP over SONET [7] (LIPE, CIP, PPP/HDLC) or IP over ATM [7] (AAL5, PPP/AAL2, CIP).

11.10.1 Node B Application Part NBAP

The NBAP protocol provides all the control functionality required between a BTS and an RNC. NABP is the communication protocol used on the Iub interface between the RNC and the Node B. Its procedures are divided into two classes—common and dedicated. Common NBAP procedures are defined as all general and non-UE specific procedures including those used to establish an initial context for the mobile user.

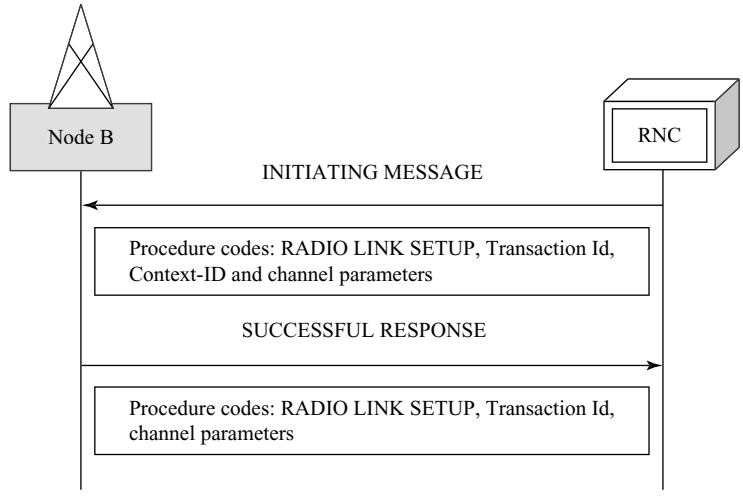


Fig. 11.20 *NABP message flow*

Dedicated NBAP procedures relate to control of the UE once an initial context has been established. The UE then manages radio links through the process of establishment, addition, reconfiguration and release of radio links; manages base station for cell configuration and scheduling of broadcast information, controls the common control channels RACH and FACH, measures the power and reports to RNC, and also does fault management by generating error signals. Fig. 11.20 shows the NABP message flow.

The NABP protocol is used to configure and manage Node B and set up channels on the Iub and Uu interfaces. The NABP **initiating message** transports the procedure request. The Elementary Procedure (EP) is the unit of interaction between the CRNC and Node B. EP is identified by the Procedure Identification Code. The CRNC Communicaion Context contains all information for the CRNC to communicate with a specific UE. The context again is identified by the Context Identifier. The EP consists of **initiating message** and a **response message.**

Procedures for NBAP are:

1. Radio link establishment
2. Common channel establishment
3. Call set-up

As an example, consider the initial connection request for establishing a radio bearer as depicted in Fig. 11.21. The RNC will send radio link set-up request to the Node B. If successful, it will result in the following action on the part of Node B:

1. Node B reserves the necessary resources, as specified in the set-up request.
2. Node B begins reception on the link.
3. Node B reponds to the RNC with a radio link set-up response message.

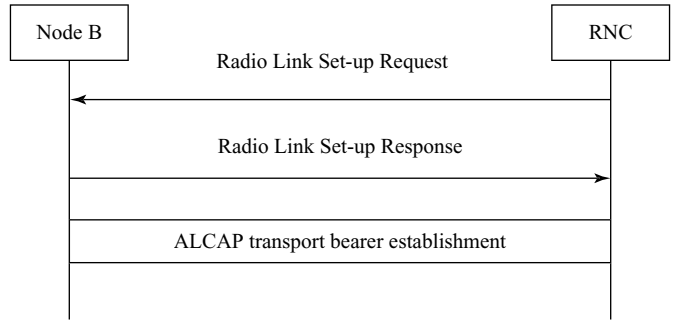


Fig. 11.21 *NBAP radio link set-up procedure*

The actual bearer service will be initiated within the transport layer on request of the RNC using transport layer signaling protocol ALCAP, such as AAL2 signaling. The NABP response provides an end-system address (AAL2 address) to Node B. A binding ID is generated by Node B and sent over NABP to the RNC. It is then passed and carried to ALCAP signaling protocol as an identifier to link the NBAP and ALCAP signaling procedure for a given transaction.

11.10.2 Radio Network Subsystem Application Part (RNSAP)

RNSAP is the communication protocol used on the Iur interface between RNCs. This procedure enables Inter-RNC soft handover. Mobile devices are connected to more than one radio link and the Node Bs are under the control of different CRNCs. Basic functions of RNSAP are mobility-related procedures within RAN such as paging and signaling traffic, procedures to handle dedicated channels on the Iur interface like radio link management and dedicated channel measurement, commom-channel procedure related to common transport RACH and FACH, and global procedures that do not connect the particular user context.

RNSAP Signaling Transfer Figure 11.22 shows the signaling transfer message for RNSAP.

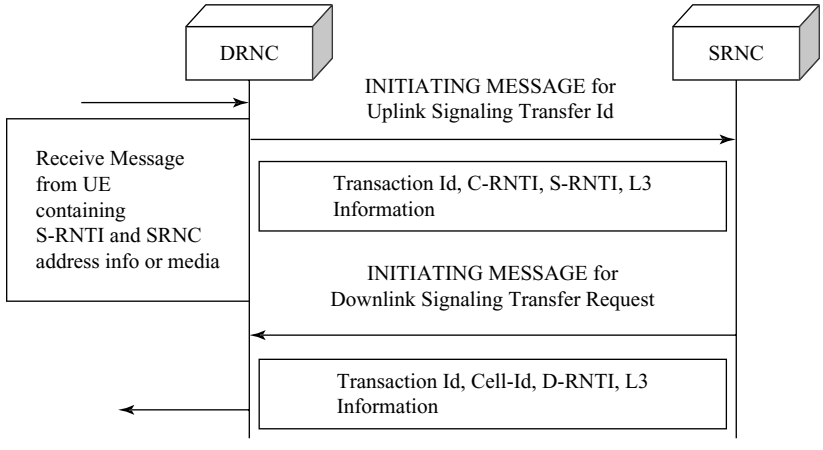


Fig. 11.22 *Procedure for RNSAP signaling transfer*

DRNC receives the radio resource control signal from the UE on the CCCH for its SRNC that contains S-RNTI and the ID of the SRNC so that DRNC can pass the message to the correct SRNC. For this purpose, it uses the uplink signaling transfer identification. To reply to the UE in the downlink channel, the SRNC will send the downlink signaling transfer request that contains a signaling message to send to the UE with the appropriate cell ID that was included in the uplink message. On receiving this message, the DRNC transfers this RRC signaling to the UE on the CCCH in the cell identified by the received cell ID.

11.10.3 Radio Access Network Application Part (RANAP)

The RANAP provides the signaling service between UTRAN and CN which is required to fullfil the RANAP functions. The RANAP services are divided into three groups on the basis of Service Access Points (SAPs)—general control services, notification services and dedicated control services. The RANAP covers procedures for both the circuit and packet domains of the core network.

The key functionalities of RANAP are Radio Access Bearer (RAB) management, establishment, modification and release of radio access bearers, SRNS relocation, i.e., shifting of radio resources and functionality between RNCs, Iu connection release, Iu load control, core network user paging services, transfer of non-access statum information (NAS) such as transfer of location and routing area update messages, transfer of security keys for ciphering and integrity protection to the radio access network and also general reporting of error situations. Figure 11.23 shows the RANAP procedures.

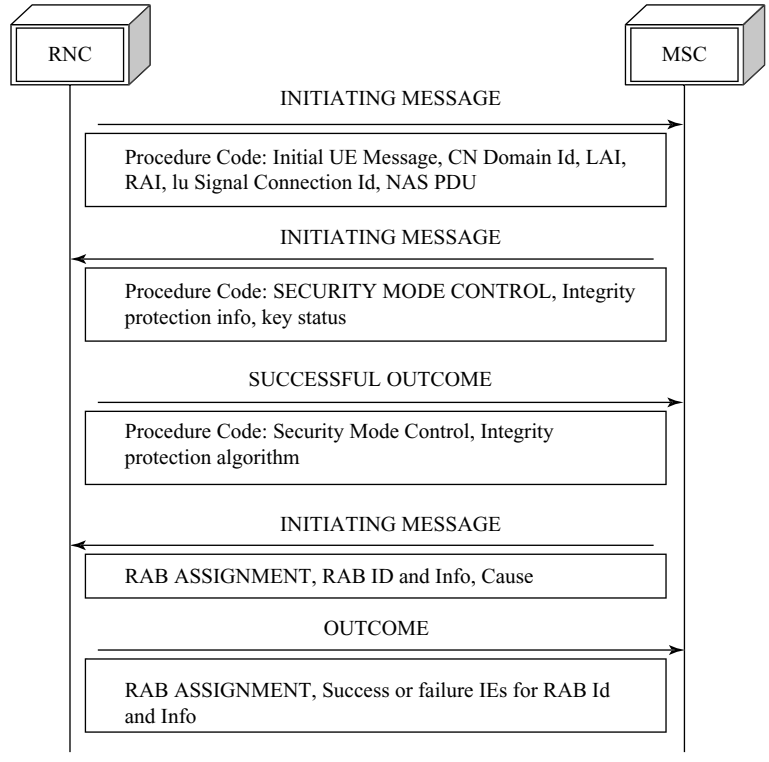


Fig. 11.23 *RANAP message flow*

11.10.4 ATM Adaptation Layer Type 2–Layer 3 AAL2L3 Protocol

For UMTS networks, the AAL2L3 represents the ALCAP functionality which is a generic name for the transport layer signaling protocol. It is used for the set-up and release of transport bearers. The AAL2 signaling protocol provides the signaling capability to establish, release and maintain AAL2 point-to-point connection across the ATM layer. User plane transport bearers for Iur interface are established and released by the AAL2L3 in the serving RNC. The binding identifier is obtained and tied to a radio application procedure when a first AAL2L3 message is received over Iur interface in the DRNC. For Iub interface, user transport bearers are established and released by AAL2L3 in the CRNC.

11.11 RAB ESTABLISHMENT, MODIFICATION AND RELEASE

RAB set-up between the RNC and core network (CN) is required whenever a user wishes to establish a call. The CN establishes radio access bearer by sending an RAB assignment request that contains the necessary QoS parameters for the RNC to determine the requisite resource allocation for the radio link. The RAB assignment request message initiates the establishment of a radio link and its associated radio bearers transport over the Iub/Uu interface. It contains the RAB identification, RAB parameters (such as traffic class, maximum bit rate, guranteed bit rate, maximum SDU size and parameters, etc.), user plane information (transparent or support for predefined SDUs), transport layer information (AAL2 ATM address of the CS-CN element and IP address for PS-CN, binding Id with the AAL2 signaling for CS-CN and GTP tunnel identifier for the PS-CN). Upon receiving the RAB assignment request for CS connection, the RNC invokes transport layer signaling AAL2 to establish the AAL2 bearer for the RAB. For PS connection, the necessary information is sent within the information request.

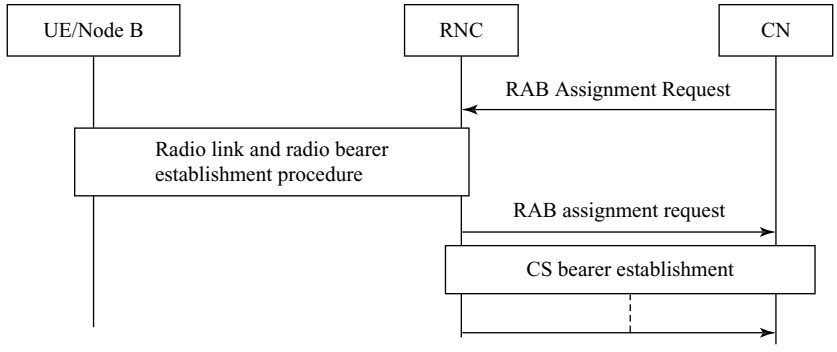


Fig. 11.24 *RAB set-up procedure*

For successful connection, the RNC replies with RAB assignment response message containing the RAB ID unique to the request message. The CN initiates the Iu release command in case of no RAB connection establishment on the Iu interface. On receipt of this request, the RNC will clear all the allocated resources and reply with Iu release message.

11.12 MOBILITY MANAGEMENT FOR UMTS NETWORK

In order to track the UEs, cells (Node Bs) in the UMTS service area are partitioned into several groups. Generally, in the CS domain, cells are partitioned into Location Areas (LAs) whereas in PS domain, cells are partitioned into routing areas (RAs) which is a subset of LA. The RA is tracked by SGSN during a PS connection. In UMTS, the cells are further partitioned into several UTRAN RAs (URAs). The URA and the cell of an UE are tracked by the UTRAN. Tracking of UTRAN is triggered by the establishment of RRC connection between the UE and UTRAN.

The operation of packet core consists of three states, referred as the packet mobility management (PMM) states. These Iu states are PMM-detached, PMM-idle and PMM-connected. In order to establish a mobility-management context, the mobile device has to perform a GPRS attach procedure. The packet-switched signaling connection consists of the Radio Resource Control (RRC) part and Iu connection.

Mobility Management (MM) functions for PS-based services are described below:

11.12.1 PMM–attach Procedure

This procedure allows the PS service domain of the network to be known. The PMM-attach procedure is required to be executed after the power of UE is on in order to get access to the network PS services. The location of UE is known in the SGSN to the serving RNC. Unlike the GPRS system, the position of a mobile user is known up to the cell level by the SGSN. Tracking is done by the SRNC up to the URA level. A packet-switched signaling connection is established between the UE and the SGSN during the PMM-connected state.

11.12.2 PMM-detach Procedure

This allows the UE or the network to inform each other that the UE will not access the SGSN-based services. The UE is not known to UMTS PS service and is not reachable to the network. In this situation, the UE may perform the attach procedure.

11.12.3 PMM–idle Procedure

In this state, the location of UE is known in the SGSN to the accuracy of an RA. At this state, paging is required to know the position of the UE at the cell level. The mobile device performs routing area update if its RA changes.

Figure 11.25 shows the mobility management functionality steps for a UMTS network.

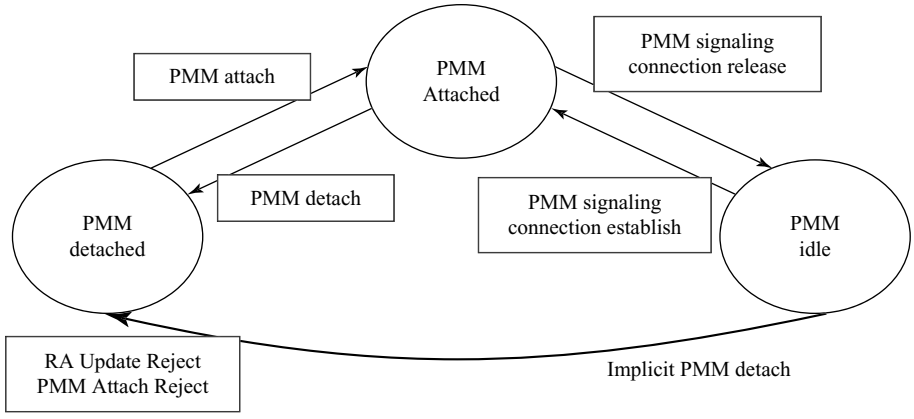


Fig. 11.25 *Mobility Management for UMTS PS service*

11.13 UMTS SECURITY PROCEDURE

The security procedure includes authentication, user identity confidentiality, such as P-TMSI reallocation and signature. In a UMTS, mutual authentication both for a user and the network is possible. Only an authorized user can access the network and users can know that the network they are connecting to is the valid one. When a new user attempts to access the network or to perform location update, the VLR/SGSN will send an authetication request to the HLR/AuC. The IMSI is used to identify the user, and HLR will then use the user key *K* to generate a set of authentication vectors to return to the VLR/SGSN. The authentication vectors are random number RAND, expected response XRES, cipher key CK, integrity key IK and authentication token AUTN. For a UMTS user, the USIM uses ATUN value to validate the network and generates its result (RES) using the RAND and its keys K, CK, IK. The RES is sent back to VLR/SGSN and is compared with XRES. If matching is obtained then the user becomes an authenticated one. USIM and VLR/SGSN will distribute these keys to the relevant units to perform encryption and integrity. The serving RNC is the relevant unit for this function. After the authentication procedure, the VLR/SGSN decides which integrity and encryption algorithms are to be used (UIA/UEA) in order of preference. The decision is then issued to SRNC with the respective keys IK and CK in the RANAP security mode command. The SRNC compares the selected UIA/UEAs required to be compatible for the UE and selects the most high-priority match. The SRNC generates a random value FRESH which is valid for the duration of signaling connection. The UE is then informed of this choice through the RRC security mode commands containing this FRESH. The message authentication code for integrity MAC-I is attached with this FRESH. This enables the UE to validate that message coming from the authenticated source. On acceptance, the UE sends a response to SRNC by sending the RRC security mode complete which (SRNC) again verifies the message and sends RANAP security mode complete message to the core network. Thus, a secured relationship is established between the UE and the RNC where all the control and data can be transferred by checking its integrity and encryption.

The UMTS performs encryption using a stream cipher and the encryption algorithm generates a keystream which is added bit by bit to the plain text to generate a cipher text. Confidentiality protection may be applied to both data and signaling messages.

11.14 UMTS HANDOVER

One of the most important aspects of the WCDMA system is the ability to handle soft handover process where a new connection is established before the existing connection is teared off. The soft-handover procedure maintains signal quality by applying diversity at the minimum power level [7]. In soft handover, the

UE is connected to more than one Node B (BTSs). All the Node Bs will by default transmit 'Transmit Power Command' messages. The rule to follow is 'lesser the transmission power, the better'. This would enable the UE to decrease the transmission power even if it loses contact with some Node Bs. An alternative method is Site Selection Diversity Transmission (SSDT). The RNC mesures the actual radio interfaces connected to all Node Bs to a particular UE using UTRAN option. It decides that some Node Bs should stop transmission of DCHs and TPC command, leaving only the best radio contact that will be the UE server in the downlink.

Now consider a case when a mobile user establishes the call, moves towards an adjecent cell employing the same frequency and performs an intra-RNC soft handover, as is shown in Fig. 11.26.

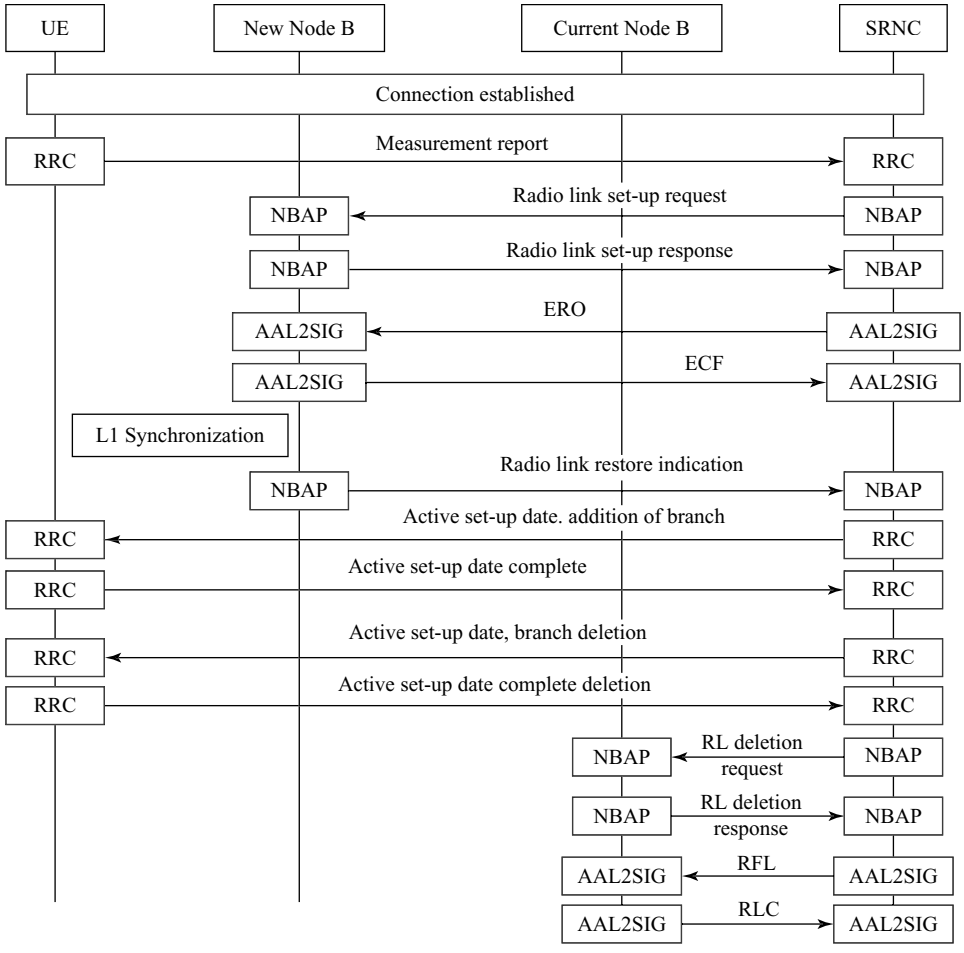


Fig. 11.26 *Intra RNC soft handover*

11.14.1 Intra-RNC Soft Handover Process

1. The measurement report is sent to the SRNC by the UE after it fullfills the criteria from the neigboring cell.
2. The SRNC performs an evaluation of the report and decides to include the new cell to the active set. The SRNC starts the radio link set-up procedure with the new Node B. For soft handover, the required parameters are passed to the same Node B with a new cell.

3. Then AAL2 bearer service is established across the Iub interface for this branch.
4. Achieving the L1 synchronization with the UE, the new Node B sends a radio-link restore indication message to the SRNC.
5. In reply, the SRNC informs the UE that the cell that was under monitoring is to be placed at an active-set with an active-set update message. The UE sends a reply with the active-set update complete. The UE is now in soft handover process and may use two active radio connections.
6. If the previous Node B does not contribute much then SRNC can decide to remove the cell from the active set by a measurement report. The SRNC once sends the active-set update message to the UE to remove the first radio link from the active set.
7. The SRNC requests the Node B to delete the radio link by sending a message radio link request/ radio-link deletion response.
8. RNC releases the AAL2 connection with REL/RLC.

Inter-RNC soft handover is required when the Iu interface needs to be relocated from one RNC to another or possibly from one CN element to other. This situation occurs in order to support roaming of a moble device between geographical areas controlled by different RNCs/ different SGSN/MSCs. Relocation may be intra/inter SGSN, intra/inter MSCs and also inter RNCs. In every relocation, the role of the serving RNC is transferred to the target RNC, called drift RNC.

Summary

This chapter describes the detailed architecture and operation of the UMTS network with the description of protocols and signaling procedures. The UMTS network is a vast subject and only an overview of UMTS for logical, transport and physical channels with emphasis on some important protocols like NBAP, RNSAP, RANAP are explained here. To explore the application of ATM to a UMTS network, an understanding of the structure of ATM protocol is required, some basics of which are provided in this chapter. For further study, readers can follow the references provided in this chapter.

References

[1] Lu, W.W., *Broadband Wireless Mobile, 3G and Beyond*, John Wiley, 2002.
[2] *3GPP, Technical Specification Group, Radio Access Network, 25.401 v3.2.0, UTRAN overall description (Release 1999)*, March 2000.
[3] Rapeli, J., *UMTS Targets, System Concept and Standardization in a Global Framework*, IEEE Personal Communication, pp 20–28, Feb 1995.
[4] *3GPP, Technical Specification Group, Radio Access Network, 25.202 v3.4.0, Services Provided by the Physical Layer (Release 1999)*, March 2000.
[5] *ETSI, Overall Requirements on the Radio Interface of the UMTS, ETSI TR 04-01/SMG-050401, Technical Report*, European Telecommunications Standards Institute, April 1998.
[6] Halonen, T., J. Romero, and J. Melero, *GSM, GPRS and EDGE Performance–Evolution Towards 3G/ UMTS*, John Wiley, Chichester, 2002.
[7] Bannister, J., P. Mather, and S. Coope, *Convergence Technologies for 3G Networks, IP, UMTS, EGPRS and ATM*, John Wiley, 2004.
[8] Dixit, S., Y. Guo, and Z. Antoniou, '*Resource Management and quality of service in Third Generation Wireless Networks*', IEEE Communication Magazine, 39, pp.125–133, 2001.
[9] *WCDMA for UMTS—Radio Access for Third Generation Mobile Communication*, edited by H Homa and A Toskala, New York, Wiley, 2000.
[10] *UMTS overview: UMTSWorld.com*, 1999–2003.

Questions for Self-Test

11.1 The first version of UMTS architecture (R99) is based on
a. WCDMA technology b. CDMA2000

11.2 UMTS development is under
a. 3GPP b. 3GPP2

11.3 Part of GSM/GPRS network can still be in used in R99 UMTS networks.
a. True b. False

11.4 The idea behind R4 UMTS networks is to integrate circuit-switched networks and packet-switched networks into a single entity.
a. True b. False

11.5 UMTS release 5 (R5) has a control layer that is responsible for handling the signaling for multimedia sessions.
a. True b. False

11.6 To access a UMTS network, a customer needs to have
a. new handset for 3G services with new SIM (USIM)
b. can use GSM/GPRS handset

11.7 The main difference between the GSM SIM and USIM is that
a. USIM is downloadable and can be accessed via air interface and modified by the network
b. it can store JAVA application
c. higher capacity than GSM SIM
d. all the above

11.8 The mobile terminal for UMTS can be single or multi-radio mode.
a. False b. True

11.9 Node B can handle higher data rate than GSM BTS.
a. False b. True

11.10 All decision-making processes for connection in UMTS are done at
a. RNC b. Node B c. SGSN

11.11 Within UTRAN, mobility is managed by
a. SRNS b. DRNC c. CRNC

11.12 UMTS networks evolved from _____.

11.13 The UMTS air interface is based on _____.

11.14 The idea of release 4 (R4) is to merge the _____ and _____ network into a single entity.

11.15 The IP Multimedia subsystem is responsible for handling _____.

11.16 DRNC handles UEs that are handed over from _____.

11.17 _____ provides radio resources to SRNC to allow soft handover.

11.18 RACH is an _____ transport channel.

11.19 Physical channel is the actual channel on the air interfaces defined by _____ and _____.

11.20 Iur interface works between _____.

11.21 The _____ interface works between user and network.

11.22 _____ is the intended protocol to be used in the IMS services.

11.23 The new integrated database similar to HLR in IMS based R5 is called _____.

11.24 Describe the UMTS FDD and TDD mode of operation.

11.25 Describe the UMTS logical and physical channel structure.

11.26 How are user and control plane protocols separated in a UMTS architecture?

11.27 What is the transport channel? What is its main function?
11.28 Illustrate and narrate the UMTS time-slot allocations for uplink and downlink dedicated physical channels.
11.29 What are the new interfaces required for the UMTS R4 network? Mention their functions.
11.30 Highlight the purpose of IMS use in UMTS R5. What are the two main components of IMS?
11.31 What are the two important gateways introduced in R4? What are their functions?
11.32 Why is ATM used in physical and data link layer used in UMTS network?
11.33 Describe the UMTS R5 revolutionary steps.
11.34 Describe the key functionalities of RANAP.
11.35 What are the main functions of UMTS radio access bearer service?
11.36 With respect to a circuit diagram, describe the three states of mobility management procedures in UMTS networks.
11.37 How is security maintained in UMTS services?
11.38 How many types of handoff are possible in UMTS networks? Describe the basic handoff procedures when mobile terminals move within the RNC.
11.39 Describe RAB Establishment, Modification and Release procedures in UMTS networks.
11.40 What was the idea of establishing UMTS R4?
11.41 What are the three different states of UMTS mobility management? Explain with respect to pictorial representation.

Overview of Internet Protocol and Mobile Internet Protocol

12

Introduction

The use of a logical architecture with Internet Protocol (IP) address allows a flexible routing procedure that always finds the best route between two end-to-end partners. The convergence of fixed, mobile and IP-based data networks offers new possibilities for network operators to make more efficient use of their networks; and also for end users, who tend to profit from new applications and a more transparent billing system. The operators can reduce the costs for network management and for the transmission of data, because now only one network architecture would be administered. The combination of different systems in one solution allows for the development of new applications, making use of the interaction of voice calls and Internet connections at the same time within the same application. A common IP core network will support multiple types of radio technologies.

IP-based networks are more suitable for supporting the rapidly growing mobile and multimedia services. IP-based protocols are independent of the underlying technology and are better suited for the seamless services over different radio technologies. IP-based wireless networks introduce many new technical challenges like network architecture, signaling, mobility management, security and quality of services. In the development process of wireless networks, evolutionary approaches are taken that are used to migrate wireless networks to full IP-based networks. Evolution starts in the core networks as 3G core networks are based on IP networks.

The increasing variety of wireless devices offering IP connectivity, such as PDAs and handheld and digital cellular phones, is beginning to change our perceptions about the Internet. The Internet offers access to information sources worldwide, but IP does not support user mobility. The Internet Engineering Task Force (IETF) has been standardising IP-based protocols for enabling mobile Internet that is designed to work over any radio system.

Mobile IP is the IETF standard that hides the IP address of the roaming users while they are moving from one network to another by providing a temporary 'care' of address in foreign networks. The Mobile IP Working Group has developed routing support to permit IP nodes (hosts and routers) using either IPv4 or IPv6 to seamlessly 'roam' among IP subnetworks and media types. The Mobile IP method supports transparency above the IP layer, including the maintenance of active TCP connections and UDP port bindings.

This chapter deals with the fundamentals of IP and mobile-IP operations which are the building blocks of mobile computing and networking. In mobile networking, computing activities should not disrupt when the user changes the computer's point of attachment to the Internet as all the needed reconnection occurs automatically and non-interactively.

12.1 BRIEF OVERVIEW OF INTERNET PROTOCOL

The Internet protocols are the world's most popular open-system protocol suite under the IETF. They can be used to communicate across any set of interconnected networks and are equally well suited for LAN and WAN communications. The standards are published through a process of Internet drafts, Request for Comment (RFC) documents and standard documents. The Internet protocols consist of a suite of communication protocols, of which the two best known are the Transmission Control Protocol (TCP) and the Internet Protocol (IP).

IP is a Layer3 (Network Layer) protocol in the Internet Protocol suite. Packets are routed with the help of addressing information and some control information. RFC 791 is the documentation of IP, which is the heart of the Internet protocols. The two main functions of IP are as follows.

1. Connectionless best effort delivery of data to the destination address, and
2. fragmentation and reassembly of datagrams to support data links with different maximum transmission units.

12.1.1 IP Packet Format

An IP packet contains several types of information, as illustrated in Fig. 12.1. There are 14 fields, each having a distinct function and meaning within the IP packet format.

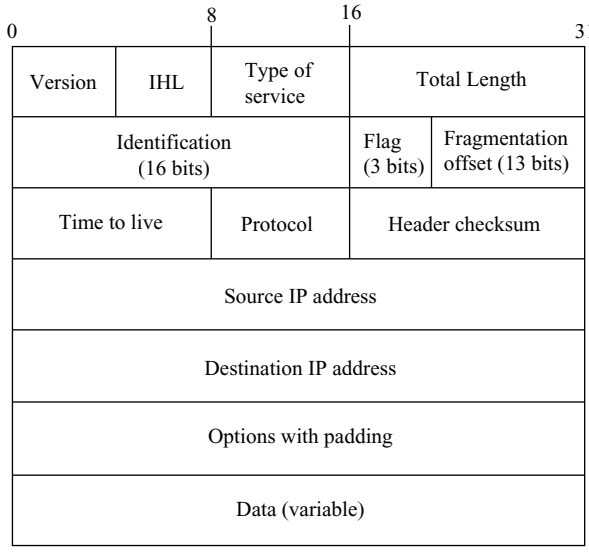


Fig. 12.1 *IPV4 packet format (32 bits)*

1. **Version:** Which version the IPv4 or IPv6 is currently using.
2. **IHL-IP:** Header length indicates the datagram length in 32-bit words.
3. **Type-of-Service:** It specifies the way the upper-layer protocol would like a current datagram to be handled. Various levels of importance are assigned. The types of fields specify reliability, precedence, delay, and throughput parameters.
4. **Total length:** This specifies the length data and header together in bytes of the entire IP packet before fragmentation.
5. **Identification:** It contains an integer that identi-fies the current datagram. This field is used to help piece together datagram fragments. When fragmenting the IP packet, an identification field is used as the same value for each fragment of the original datagram.
6. **Flags:** It consists of a 3-bit field of which the two low-order (least-significant) bits control fragmentation. The low-order bit specifies whether the packet can be fragmented. The middle bit specifies whether the packet is the last fragment in a series of fragmented packets. The third or high-order bit is not used.
7. **Fragment offset:** It indicates the position of the fragment's data relative to the beginning of the data in the original datagram. This field is used to help to put the fragment back together in the correct position at the destination.
8. **Time-to-Live:** It maintains a counter that gradually decrements down to zero at the point of a discarded datagram, thus helping in preventing looping for any misrouted packet.
9. **Protocol:** It indicates the higher-layer protocols like TCP or UDP that receive incoming packets after IP processing is complete.
10. **Header checksum:** It is required for the error correction in a header. It protects against corruption of packets and discards packets with an incorrect header checksum.
11. **Source address:** It specifies the address of the sending node.
12. **Destination address:** It specifies the address of the receiving node.
13. **Options:** It allows IP to support various options, such as security.
14. **Data:** It contains upper-layer information.

Due to the ever-growing demand for Internet services, the Internet community believes that the IPv4 address alone will not suffice the address space required in future in order to connect thousands of routers and millions of users over the world. Many RFCs are being developed to provide expanded address space in addition to simplified header format, flow control, authentication and security. Some of the benefits of IPv6 routers are greater addressing space, built-in QoS, and better routing performance and services.

For IPv6, a 128-bit addressing scheme is used instead of 32-bit addressing. Figure 12.2 shows the IPv6 packet header format.

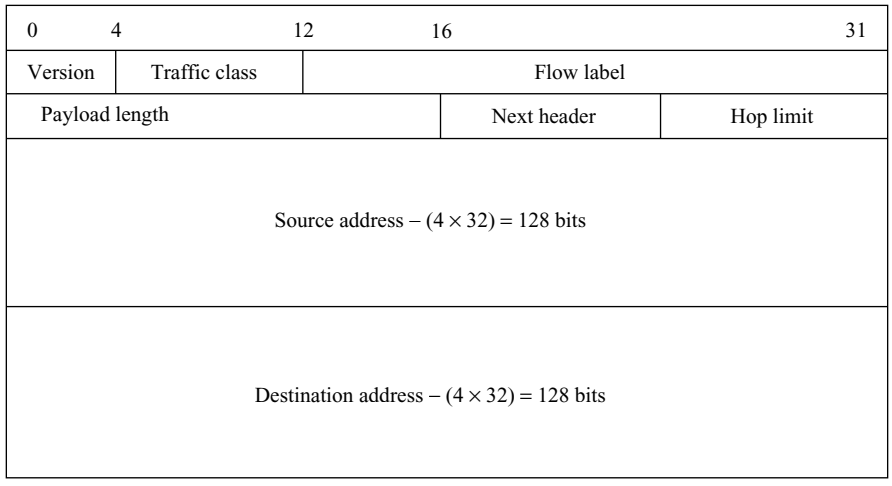


Fig. 12.2 *IPv6 header format*

The IPv6 header is much simpler than IPv4 and is of a standard 40-byte length. Hence, there is no header length field. Extension headers replace the need for fragment offset, identification and flags. Also, the checksum field is completely removed and error checking is duplicated at other layers of the protocol stack. The IPv6 header consists of the following fields:

1. **4-bit IP version (number 6):** This field contains the binary value 0110.
2. **8–bit traffic class:** It is used to classify the priorities of the IPv6 packets, and accordingly forwards them to the routers. It is equivalent to the type of service in the IPv4 header.
3. **20–bit flow label:** It may be used by a source node to label the sequence the packets for which it requests a special type of handling by the IPv6 routers (non-default Qos, real time service, etc). A flow is uniquely identified by the combination of the source address and a non-zero flow label. The flow label is assigned to a flow by a flow source node. Packets having the same flow label, source address and destination address belong to the same flow class.
4. **16–bit payload length:** It indicates the length of the payload in bytes. Any header extension is a part of the length of the payload. For jumbo payloads (65535 bytes), the payload is set as zero and the actual length of the payload is carried in the jumbo payload by the hop-by-hop option.
5. **8–bit selector-next header:** It indicates the type of header that the IPv6 header follows immediately such as ICMPv4/ ICMPv6, TCP, UDP.
6. **8–bit unsigned integer hop limit:** The hop limit is decremented one by one after forwarding the packet by each node. When zero, packets are discarded and an error message is returned. This is equivalent to the time-to-live field of the IPv4 header. For IPv6, this field allows 256 routers to traverse between the source and destination. Note that the value is measured in hops and not in seconds.

8. **128–bit source address:** It indicateds the origin of the IPv6 packets.
9. **128–bits destination address:** It indicates which packets are destined.

The most distinct feature of IPv6 packets is the use of optional headers. An IPv6 packet with a payload may consist of zero, one or more extension headers.

12.1.2 IP Class and Addressing

An IP address is a logical address at the network layer of the TCP/IP protocol suite. An IPv4 address is a unique 32-bit long address universally defined for the connection of a device such as any computer, or to any router within the Internet domain. An IP address is unique in the sense that two devices on the Internet cannot have the same IP address; otherwise an IP conflict will result. It is universal in the sense that the addressing pattern is acceptable by any host (computer) on the Internet worldwide.

As IPv4 uses a 32-bit address, its address space is 2^{32} or 4294967299. But due to the address schemes and restrictions imposed, the actual number of devices that can be connected to the Internet at a given point of time is much less than this number. In this section, the address schemes for IPv4 and IPv6 will be discussed.

IPv4 Address Format The 32-bit address is grouped into eight bits at a time separated by dots and represented in decimal format, known as dotted decimal notation. Each bit in the octet has a binary weight 1,2,4,8,16,32,64,128.The minimum value for an octet is 0, and the maximum value for an octet is 255.

Example of dotted decimal notation: 192.168.25.20, 128.45.60.2

IPv4 Address Classes There are five address classes A, B, C, D and E for IPv4. If the address is given in binary notation then the first few bits determine the class. If the address is in dotted decimal then the first byte tells the address class. The class of address can be determined easily by examining the first octet of the address and mapping that value to a class range. Figure 12.4 describes the class type addressing for IPv4.

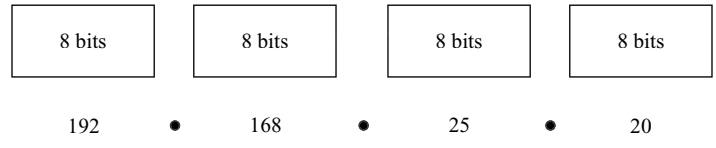


Fig. 12.3 *Dotted decimal notation*

Class A is designed for large organisations with a large number of hosts, whereas Class B is for medium-size organisations, and Class C is designed for small organisations with a small number of attached hosts. Classes A, B and C are for unicast routing, Class D is for multicast and Class E is reserved for future applications.

For IP classful addressing, the total address bytes are divided into *netid* and *hostid*. The shaded parts in Fig. 12.4 is the indicator for netid in classes A, B and C. This classification is not applied for classes D and E.

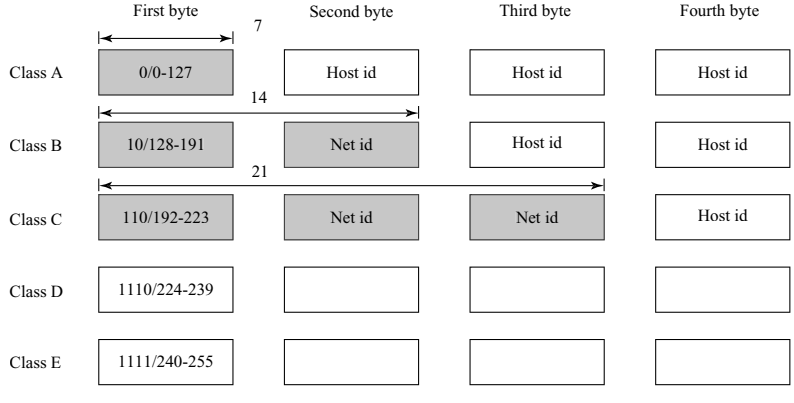


Fig. 12.4 *IPv4 address class*

The main problem of classful addressing is the possibility of wasting address space. As an example, for class A, the maximum number of hosts is $2^{24} - 2 = 16{,}777{,}214$ (one address is reserved for the broadcast and one address is reserved for the network). In many organisations there is a chance of wasting such a large number of hostids.

IP Subnet Addressing Subnetting is used for efficient use of network address. IP networks can be divided into smaller networks called *subnetworks*, (subnets). Subnetting is the division of a single network with many hosts into smaller subnetworks with fewer hosts. Subnetting also provides the network for broadcast traffic within the subnets that need not cross a router. Subnets are under a local organization's administrative control. Outside the organization, it will be seen as a single network, as the outside world has no knowledge of the organizational structure.

Let the IP address for Class B in Fig. 12.5 be 128.10.10.0. Then 128.10.11.0, 128.10.12.0 ……128.10.240.0 are the subnet addresses within the network address 128.10.10.0.

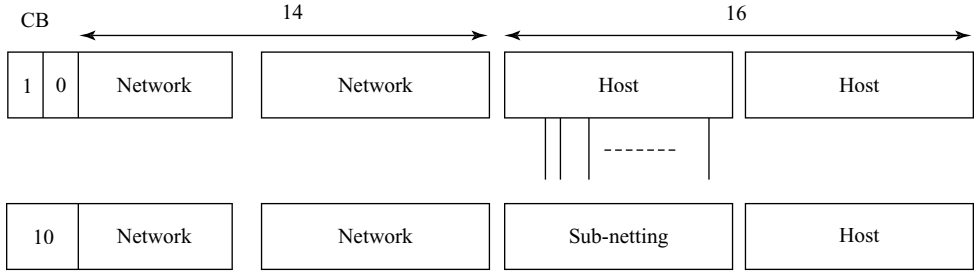


Fig. 12.5 *Class B IP addressing*

IP Subnet Mask Borrowing bits from the host field creates a subnet address. The number of borrowed bits varies and is specified by the subnet mask. Although the netid and hostid are predetermined for classful addressing, default masking can still be used which is a 32-bit number made of contiguous 1s and 0s. The mask helps to determine the netid and hostid. Figure 12.6 shows the subnet masking for classes A, B and C IP addressing.

Class	Dotted decimal	Binary representation			
A	255.0.0.0	11111111	00000000	00000000	00000000
B	255.255.0.0	11111111	11111111	00000000	00000000
C	255.255.255.0	11111111	11111111	11111111	00000000

Fig. 12.6 *Subnet masking*

The subnet mask has binary 1s in all bits specifying the network and subnetwork fields, and binary 0s in all bits specifying the host field. As a sample subnet mask for Class B, let the subnet mask be 255.255.240.0. The binary representation is given as:

11111111	11111111	11110000	00000000
Network	Network	Host	Host

Subnet mask bits should come from the leftmost bits of the host field, as illustrated above. If there are four masked bits, the possible number of subnets and hosts are given as follows:

$$\text{Subnets} = 2^{\text{number of masked bits}} - 2, \text{ Hosts} = 2^{\text{number of unmasked bits}} - 2$$

For the above example, the number of subnets and hosts are 14 and 4094 respectively. The number 2 is subtracted because one address is for the broadcast and one address is reserved for the network.

Subnet masks are used to determine the network number. First, the router extracts the IP destination address from the incoming packet and retrieves the internal subnet mask. It then performs a logical AND operation to obtain the network number. This causes the host portion of the IP destination address to be removed, while the destination network number remains.

Example

Destination IP Address	10001101	00011000	11000110	00110011
Mask	11111111	11111111	11110000	00000000
Sub-Network Address	10001101	00011000	11000000	00000000

12.1.3 IPv6 Addressing

IPv6 address bits are 128-bit long and are represented by 16-bit hexadecimal numbers as specified in RFC 2373. An example for IPv6 address is

EFDC**:**A980**:**3456**:**DEFC:C789:3255:3A70:908A

Each of the hexadecimal numbers is separated by a colon. Each part is a 16-bit number and is eight parts long, providing a 128-bit address length ($16 \times 8 = 128$). If a long string of zeros appears in an address, a double colon '**::**' is used to indicate the presence of a multiple group of zeros and can be used once. This is illustrated below:

DECC**:**0000**:**0000**:**0000**:**0000**:**3425**:**A290**:**000A

DECC**::**3425:A290:000A

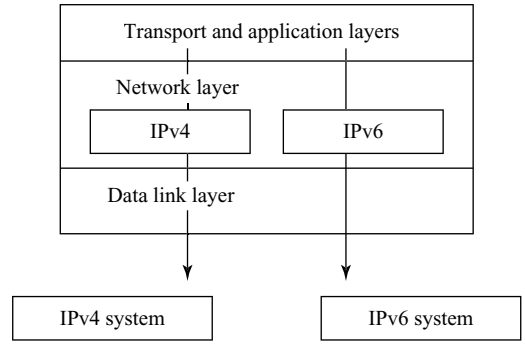


Fig. 12.7 *Dual stack IP*

IPv4 and IPv6 Address Transition There are three methods of transition strategies from IPv4 to IPv6:

1. Dual stack
2. Tunneling
3. Header Translation

Before migrating to IPv6, it is recommended that the entire host should contain both the IPv4 and IPv6 dual protocol stack.

12.2 TRANSMISSION CONTROL PROTOCOL (TCP)

The TCP provides reliable transmission of data between the two communicating hosts. TCP corresponds to the Layer4 (transport layer) of the OSI reference model, and is a process-to-process protocol. It is a connection-oriented service though the underlying protocols for internetworking are the unreliable IP datagram services. To the host, it seems to be circuit switched connection. TCP does this by sequencing bytes with a forwarding acknowledgment number that indicates to the destination the next byte the source expects to receive. Bytes not acknowledged within a specified time period are retransmitted. TCP is a reliable transport protocol, and the reliability mechanism of TCP allows devices to deal with lost, delayed, duplicate, or misread packets. A time-out mechanism allows devices to detect lost packets and request retransmission.

TCP offers full duplex communication in which data can flow in both directions at a time. Among the services that TCP provides are flow control, error control, and congestion control.

A packet in TCP is known as *segment*. Figure 12.8 is the TCP segment header.

TCP segment consists of a 20–60 byte header and are followed by data from the application program.

16-bit source port defines the port numbers of the application program, whereas the destination port address defines the port number of the application program in the host that is receiving the segment.

Sequence number is the number assigned for the first byte of data contained in the segment. This number is used to make sure that data is received in the correct order. During the connection establishment process, an Initial Sequence Number (ISN) is generated randomly by each party that differs in two directions.

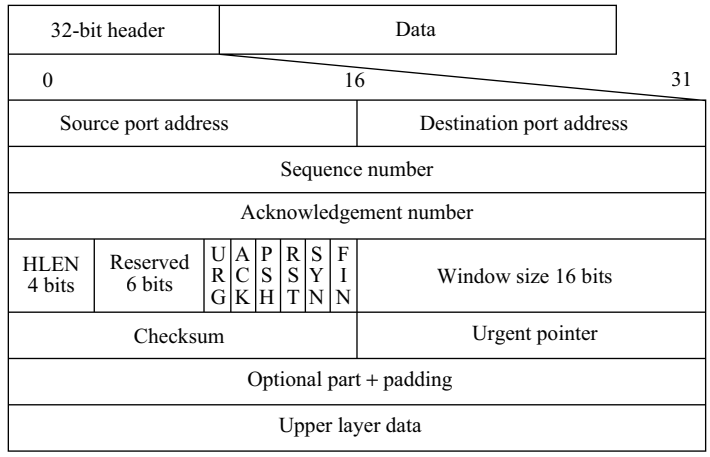


Fig. 12.8 *TCP segment format*

The acknowledgement number defines the byte number that the receiver of the segment is expecting to receive from the other communicating party. When packets are received at the destination, an acknowledge-ment is sent back to the source. If the acknowledgement is not received within a specified period, the packet is resent again. The acknowledgement number is used to match each acknowledgement with a particular ISN. This is explained in Fig. 12.9.

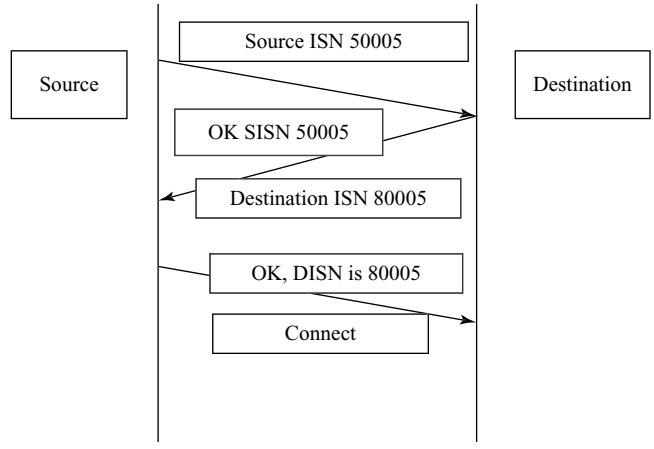


Fig. 12.9 *Three-way handshaking*

As TCP is a connection oriented protocol, before sending data, a connection is made with the remote host by a three-way handshake process as given in Fig. 12.9. The purpose of handshaking is to make sure that each end is ready to accept data and to let know the sequence number each end expects to receive. After the connection establishment, data is transferred in both directions. The TCP port numbers are used to route the data in a particular application program. This is required because there may be several applications like email, file transfer, Web browsing, etc., for one IP connection,

There are six different control bits, each with a specific meaning. These are used for flow control and connection management.

URG The value of urgent pointer is valid.

ACK The acknowledgement is valid.

PSH Request for push (for data).
RST Reset the connection.
SYN Synchronize sequence numbers during connection.
FIN Terminate the connection.

The 16-bit window size field defines the size of the window in bytes (maximum size of the window is 65535 bytes). The receiver sets a window size that determines the maximum amount of data the sender can send. If the window size is zero then the receiver can stop the data flow completely. A 16-bit checksum is used for calculating errors, a process that is mandatory for TCP. The 16-bit urgent pointer is valid if the URG flag is set. Finally, the option field can be of 40 bytes of optional information in the TCP header.

12.3 USER DATAGRAM PROTOCOL (UDP)

While TCP is a reliable and connection-oriented data-transfer protocol, UDP is an unreliable and connectionless transport-layer protocol. UDP is much simpler than TCP and faster transmission is possible due to reduced header fields. A situation may arise in TCP where retransmission time becomes large producing a delay in the overall transmission. For voice and video, a little data loss may not create much problem to user perception, but delay may cause inconsistent voice and video data transfer. In such a situation, UDP can serve as a better option as UDP supports multicasting unlike TCP. Figure 12.10 shows the UDP header format.

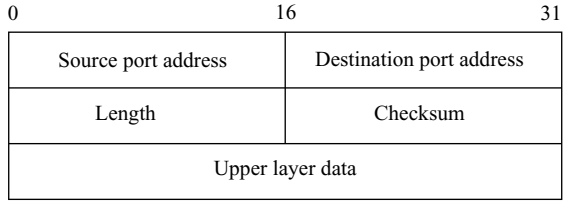


Fig. 12.10 *UDP header format*

UDP is the transport protocol for several well-known application-layer protocols, including Network File System (NFS), Simple Network Management Protocol (SNMP), and Domain Name System (DNS).

12.4 DOMAIN NAME SYSTEM (DNS)

Domain Name System (DNS) is the text name of an IP connection on the internetworking environment. This protocol provides a mapping service for the IP protocol stack. Usually, for a user it is very difficult to remember an IP address, but a user can easily remember a name when sending an email to a remote host. So we need a system that can map an address to a name. In the initial stages of the Internet, a host file that contained two columns did the mapping between the name and the IP address. Every host could maintain a host file that needed to be periodically updated from a master host file. Due to the huge size of the Internet today, this task becomes almost impossible as single host file will be too big in size to maintain. For this solution, the huge amount of data is divided into smaller parts and stored in a different computer. The host that needs mapping can communicate with the nearest computer holding the needed information. This system is named DNS. As for example, www.calcuttatelephone.com may be the name address for the Calcutta Telephone host. A user seeking connection to it will type the name into the browser. This address is then passed to a DNS server, which is responsible for translating the name into an IP address.

Domain names are arranged into a hierarchical tree name-space structure as shown in Fig. 12.11.

Each node under a tree has a label. The label can be a maximum 63-character string. The root string is a null string. A full domain name is a sequence of labels separated by dots. The last label is the root label.

As an example, let the domain name be **isical.ac.in.** This is the domain of the Statistical Institute of India administered by ISI, Calcutta. This is under the domain of ac.in, the domain allocated for all academic institutes of India.

The last character is a **(.)**, used to specify the root. It is also a fully qualified domain name as it is terminated by a null string dot. If a null string does not terminate the label then it is called a partially qualified domain name (i.e., isical).

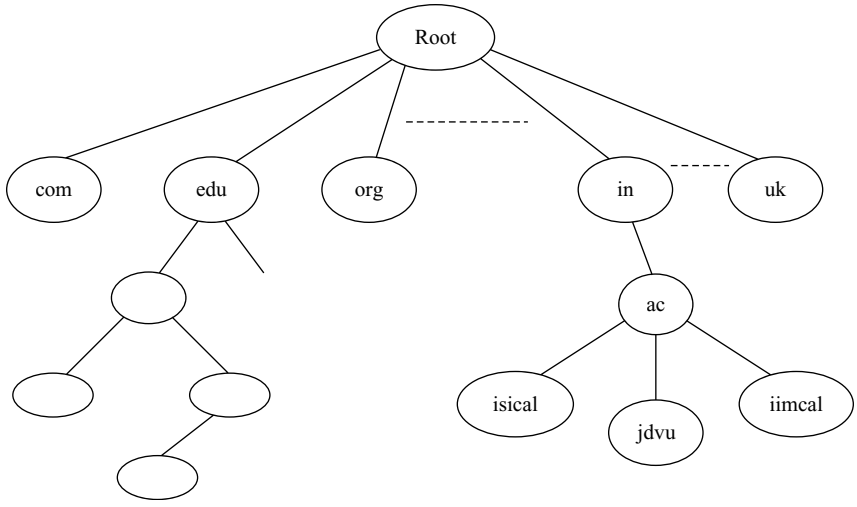


Fig. 12.11 *DNS namespace hierarchy*

A domain is the subtree of the domain name space. The name of the domain is the name of the node at the top of the subtree. Figure 12.12 shows the domain and subdomains within a domain. To distribute the information database among a number of computers called *servers*, the entire domain space is divided into many domains based on the first label like edu, com, in, uk, etc. Each of the servers is responsible for administering a large or small domain.

When mapping DNS names into IP addresses, the client machine must refer to the DNS server. In general, most client machines are configured with a default DNS server. If the local DNS server does not have the mapping for the DNS address then it will use the DNS name to work out where to look for the mapping.

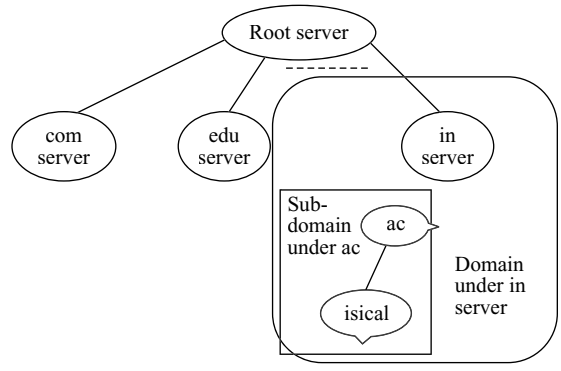


Fig. 12.12 *Division of domains*

12.5 NETWORK ADDRESS RESOLUTION PROTOCOL

Address resolution protocol (ARP) is a protocol used for mapping IP addresses (logical) to media access control (logical) MAC addresses. IP packets are encapsulated in a frame which needs a physical MAC address to deliver data to a destination node. ARP is designed for this purpose. When two machines communicate within a local area network (LAN), they must know the MAC addresses of each machine. This is done as follows. An ARP request packet is broadcast to the LAN. The request packet contains the physical and IP address of the sender and the IP address of the receiver that requires mapping. The MAC broadcast address consists of the destination address set to all 1s (F F F F F F F F F F F F). All stations on the LAN analyse this frame. Only one recognises its own IP address in the request and reply to the sender. The reply packet contains the recipient's physical and IP address. This is explained in Fig. 12.13. Figure 12.14 is the ARP packet format. It is to be mentioned that ARP works with a local cache that stores the recent copies of ARP mappings. It helps in reducing network overhead and increases performance efficiency. Before sending any ARP, the system checks its cache first to find out if any mapping exists.

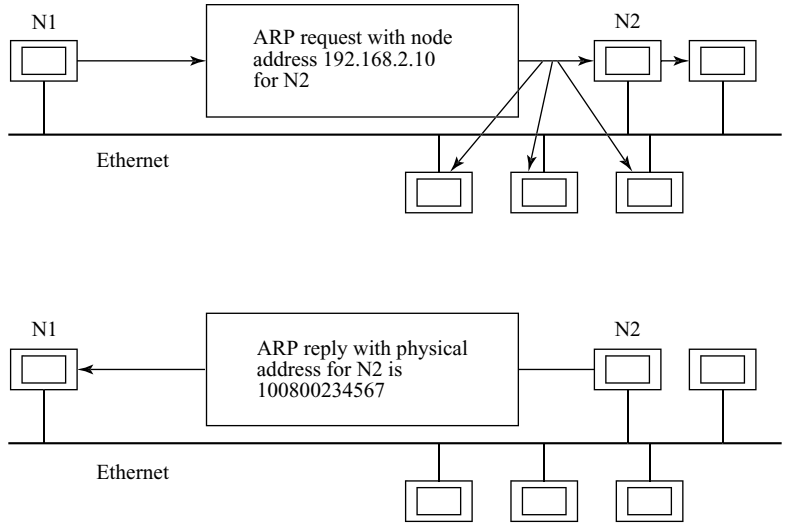


Fig. 12.13 *ARP request and reply process*

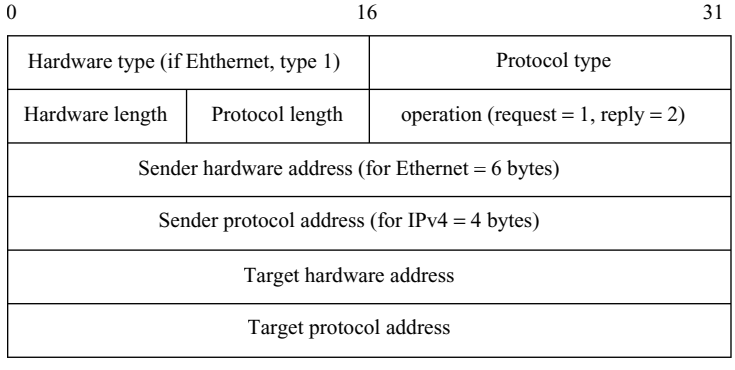


Fig. 12.14 *ARP packet format*

For more clarity the ARP request and reply procedure can be tabulated as follows.

Packet type	Destination IP address	Destination MAC address	Source IP address	Source MAC address
ARP request	192.168.2.10	FFFFFFFFFFFF	192.168.2.1	000460021234
ARP reply	192.168.2.1	000460021234	192.168.2.10	100800234567

Again, there is the Reverse Address Resolution Protocol (RARP) used to map MAC-layer addresses to IP addresses. RARP is the logical inverse of ARP and might be used by discless workstations that do not know their IP addresses when they boot. RARP relies on the presence of a RARP server with table entries of MAC-layer-to-IP address mappings. A RARP request is created and broadcast to the local network. Another machine on the same network that knows all the IP addresses will reply to the RARP request. The requesting machine runs the RARP client program, and the responding machine must be running the RARP server program.

If there is a situation where the host moves from one network to another and sometimes it needs a temporary IP address then this situation can be handled by the **Dynamic Host Configuration Protocol** (DHCP) that can provide both static and dynamic address allocation. For static allocation, the host requests a static address from

the DHCP server, which maintains database that statistically binds physical addresses to IP addresses. For dynamic allocation, the DHCP has a second database named DHCP dynamic data base that maintains a pool of IP addresses. On request, it provides the available IP addresses to the client for a specified period of time on a lease basis. When the lease expires, the client either renews the lease or stops using the assigned IP addresses.

12.6 IP ROUTING PROTOCOLS

IP routing is the way to deliver packets from the source to the destination host. When forwarding data, the routers first look at the network of the destination address. The routing table may be static or dynamic. Static means manual entries of routes by the network administrator. On the other hand, a dynamic table automatically updates whenever there is a change.

Routing protocols have been created in response to the demand for dynamic routing. IP routing protocols are dynamic. Dynamic routing calls for routes are calculated automatically at regular intervals by software in the routing devices. This is in contrast to static routing, where routers are established by the network administrator and do not change until the network administrator changes them. Static routing is suitable for simple networks. In case of dynamic routing, routers will talk to each other to exchange information based on the network topology. Using the efficient routing algorithms and topology, routers calculate the optimum route for the packets to flow to the destination. The combination of routing algorithms and the protocols needed to exchange routing information is called *routing protocols*.

IP routing specifies that IP datagrams travel through internetworks one hop at a time. The entire route is not known at the onset of the journey. Instead, at each stop, the next destination is calculated by matching the destination address within the datagram with an entry in the current node's routing table.

Today, as the size of the Internet is large, one routing protocol is not enough to handle the task of updating the routing tables of all routers. So the Internet is divided into Autonomous Systems (AS) maintained by a single network administrator.

Routing protocols are classified into two classes—internal and external routing. An example of the internal routing protocol RIP (Routing Information Protocol) is OSPF (Open Shortest Path First). These are used within an AS, managed by a single operator. Routing within an AS is also referred as *intradomain routing*. An example of external routing protocol is BGP (Border Gateway Protocol). BGP handles the routing between different ASs, to connect gateways between different Internet Service Providers (ISPs). This is also called *interdomain routing* and is illustrated in Fig. 12.15.

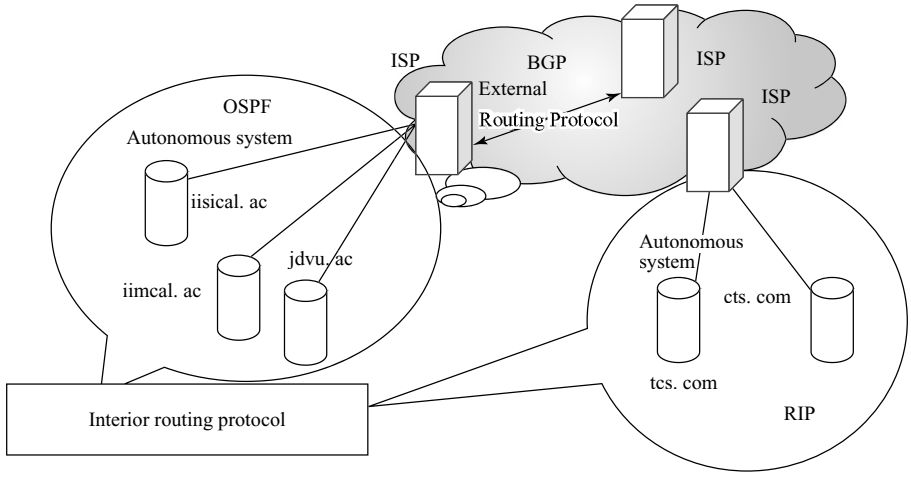


Fig. 12.15 *Internal and external routing*

Each node's involvement in the routing process is limited to forwarding packets based on internal information. The nodes do not monitor whether the packets get to their final destination, nor does IP provide for error reporting back to the source when routing anomalies occur. This task is left to another Internet protocol, the Internet Control Message Protocol (ICMP). IP also does not guarantee the order of packet delivery. The datagram between two hosts can take different paths across the network. This may be delivered at different times with variable orders. The upper layer protocol (TCP) provides the services for guaranteed and in-sequence ordered delivery.

12.6.1 Internet Control Message Protocol (ICMP)

IP is an unreliable and a connectionless best-effort delivery datagram service. It has no error control and error correcting mechanism. This is done with the help of ICMP that generates useful messages like *destination unreachable*, *echo request* and *reply*, *redirect*, *time exceeded*, and router advertisement and router solicitation. The destination unreachable, message is generated when the router cannot send the package to its final destination due to non-availability of the destination address, and thus discards the packets. On the other hand, using 'ping' command to test the reachabality generates an ICMP echo request message. When the time-to-live fields on an IP packet reach zero then the routers send ICMP time-exceeded message.

The ICMP Router-Discovery Protocol (IDRP) uses router-advertisement and router-solicitation messages to discover the addresses of routers on directly attached subnets. Each router periodically multicasts router-advertisement messages from each of its interfaces. The hosts then discover addresses of routers on directly attached subnets by listening for these messages. Hosts can also use router-solicitation messages to request immediate advertisements rather than waiting for unsolicited messages.

The IRDP offers several advantages over other methods of discovering addresses of neighboring routers. Primarily, it does not require hosts to recognize routing protocols, nor does it require manual configuration by an administrator.

12.7 BASIC MOBILE IP

Mobile IP is a proposed standard protocol that builds on the Internet protocol by making mobility transparent to applications and higher-level protocols like TCP.

IP routes the packets across the fixed-network structure, and the source and destination nodes remain within the same network. When the destination node is mobile, the new point of attachment of the mobile node is no longer within the same network. Thus, a new IP address makes transparent mobility impossible.

Mobile IP (MIP) is a standard protocol proposed by the IETF working group to solve the mobility problem. MIP uses two IP addresses, a fixed home address and a care-of-address for the new network where the mobile node presently moves on. This section will deal with the operation and functionalities of mobile IP to support a user's personal mobility. The current mobility-management protocol defined by IETF is the Mobile IPv4 (MIPv4). It first became available in RFC 1996, later it was revised in the year 2002. Due to the advantages of the IPv6 protocol over IPv4, the IETF has also defined another mobility management protocol that is based on the IPv6 platform and is known as the Mobile IPv6 protocol.

To deliver a packet to a terminal on a network, there must be an IP address by which the terminal is identified. With a fixed terminal it will work well, but when the terminal moves to a new IP network (visited or foreign), the terminal will get a new IP address from the visited network.

12.7.1 Mobile IP Types—MIPV4 and MIPv6

Next-generation wireless networks demand an architecture that would enable seamless connectivity on the move. IPv4, as such, does not provide any support for mobility. To support mobile devices, which dynamically change their point of attachment to the Internet, the IETF has standardized a protocol called MIP that alleviates the mobility problem. There are two variations of MIP—namely MIPv4 [6], based on IPv4, and MIPv6 [10], based on IPv6—which are important milestones for mobile computing. Anticipating the future

growth of the Internet, IPv6 provides relief for impending shortage of network addresses. Other advantages of the IPv6 over IPv4 are security header implementation, destination options for efficient rerouting, address auto-configuration, avoidance of the ingress filtering penalty and also quality of service capabilities. The design of MIPv6 represents a natural combination of the previous experiences, gained from the development of MIPv4, together with the opportunities provided by the design and deployment of IPv6 and the new protocol features offered by IPv6.

We can expect inexpensive, computationally powerful mobile devices running IP-based applications in the future. The ultimate goal is to provide multimedia data services on the move. To meet this, demand-efficient mobility support becomes the key focus for future Internet's performance. Thus, there is a need to pay attention to mobility support. Extensive research and standardization is on its way in the areas of MIPv4 and MIPv6.

12.7.2 Mobile IP: Concept

Confident access to the Internet anytime, anywhere will help to free us from the ties that bind us to our desktops. Mobility means to make sure that mobile wireless devices can attach to the Internet and remain attached even on the move. It was apparent that modifying the IP address would solve this problem including those of application transparency and seamless roaming.

IP mobility management comprises a unique functionality in specific places within a network that enables mobile hosts to move across different IP networks and still have uninterrupted services. We are interested in allowing mobile hosts (MH) to continue receiving services when moving to a new IP subnet. So, active TCP sessions should not be reset after a mobile node changes its access point. The source node corresponding to the MH does feel that the MH remains stationary. Keeping track of the MH's current position and to deliver packets, a special mechanism named Mobile IP is needed. Before explaining the Mobile IPv4 operations and functionalities, let us explain the Mobile IPv4 entities with the help of Fig. 12.16.

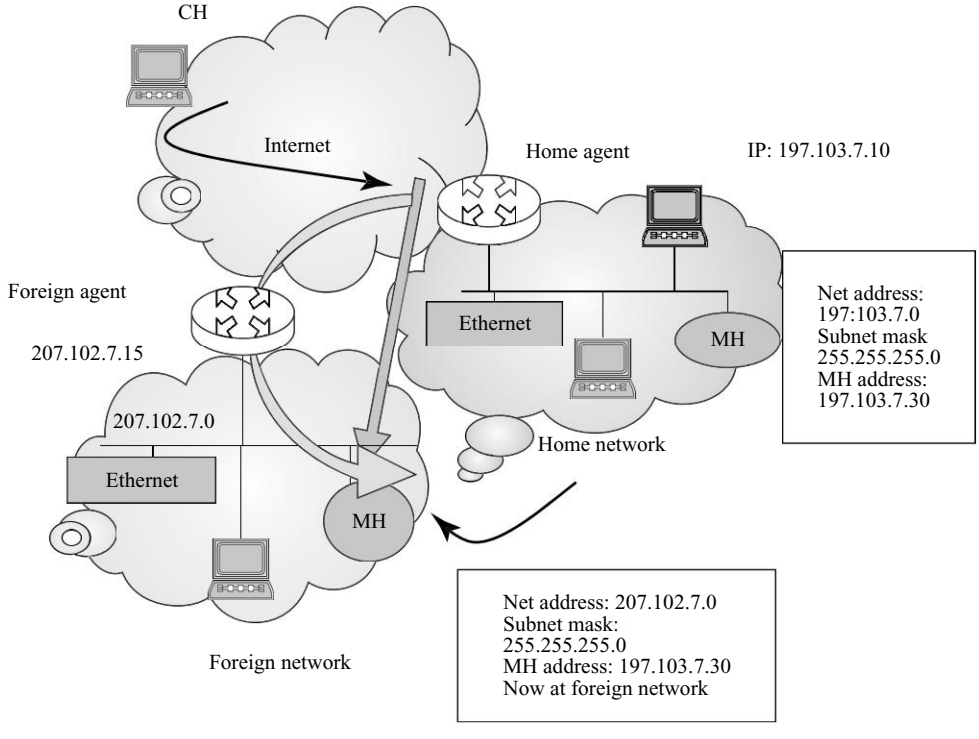


Fig. 12.16 *Mobile IPv4 functional entities*

12.7.3 Four Basic Entities for MIPv4

1. **Mobile Host (MH):** A host or router that changes its point of attachment from one network to another without changing its IP address.
2. **Correspondent Host (CH):** A peer with which the mobile node is communicating to a node, either mobile or stationary.
3. **Home Agent (HA):** A router on a mobile node's home network, which delivers datagrams to departed mobile nodes, and maintains location information for the host.
4. **Foreign Agent (FA):** A router on a mobile node's visited network which cooperates with the home agent to complete delivery of datagrams, and provides care-of-address to the MH in a foreign network.

A mobile IP operates only when the MH is away from the home network. It has two IP addresses—a fixed home address and a care-of address that changes at each new point of attachment.

The mobile host has a **Home Network (HN)** whose network address prefix matches that of the MH's home address. The home address is a permanent globally unique and routable IP address of the MH to which any correspondent host communicates regardless of the location of the MH. The **HoA** denotes the home address of the MS. When an MH is inside the HN, it receives and sends data as a fixed terminal without using Mobile IP. A router on the HN may act as a Home Agent (HA) that maintains up-to-date location information of the MH, intercepts packets addressed to the MH and tunnels this to the current location. Any visited network of the MH is called the **Foreign Network (FN).** Within the FN, the MH is identified by the temporary **care-of-address (CoA)** that is assigned by the FA of the FN, a router that provides routing services to the visited MH under the FN.

12.7.4 Mobile IPv4 Operations

The Mobile IPv4 performs the following operations.

Agent Discovery Mobility agents FA or HA advertise their availability on each link for which they provide services. For communication, the MH needs to know the agent's IP address. Generally, the MH is configured with its HA's IP address, as it is a fixed IP address. When away from home, the MH obtains a temporary care-of IP address from the FA in which it resides. The process for an MH to discover the mobility agents and receive information from the agents is called *MIPv4 agent discovery*. Mobility agents advertise their presence and services to the mobiles by sending **agent advertisement** messages that are periodically broadcast to all mobiles. An MH may not wait for long to get an agent advertisement. The agent advertisement allows for the detection of mobility agents, lists one or more available CoAs and lets MNs determine the network number and status of their link to the Internet, whether it is on its HN or FN.

A mobile node can optionally solicit an agent advertisement message from any locally attached mobility agent through an **agent solicitation message**. A mobile node receives these agent advertisements and determines whether they are on its home network or a foreign network. All mobility agents need to reply to any received solicitation.

The MIPv4 uses the Internet Control Message Protocol (ICMP) Router Discovery Message for **agent discovery** that is a standard message defined by IETF. Two types of messages are possible: (i) ICMP router advertisement message sent by the routers to the mobile hosts, and (ii) ICMP router solicitation message sent by the mobile hosts to a router to send the router advertisement message for an IP address of the router.

The **MIPv4 agent advertisement** message includes the extension information that carries MIPv4 specific information in addition to the ICMP router advertisement message. The format for an MIPv4 agent advertisement message is illustrated in Fig. 12.17 along with the header formats for MIPv4 mobility agent advertisement extension and prefix extension.

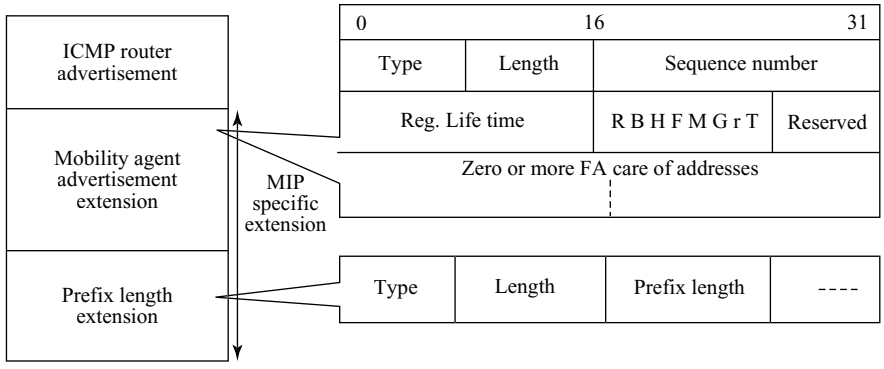


Fig. 12.17 *Mobile IPv4 agent advertisement message format*

To indicate the MIPv4 mobility agent advertisement, the **type** field value is 19. **Length** indicates the length in octets for the extension header starting from sequence number field to the end of extension. **Sequence number** is the number of the agent advertisement message. **Registration Lifetime** is the time in seconds that the agents wait to accept the registration request. There are 8 fields, each of 1 bit, within the mobility agent advertisement message having specific functional significance.

1. **R:** Indicates registration requirement with one of the FAs within the link of the foreign network.
2. **B:** Busy signal, FA will not accept any more registration request.
3. **H:** Home agent will offer services as HA on the link of the agent advertisement message sent.
4. **F:** Foreign agent will offer services as an FA on the link of the agent advertisement message sent.
5. **M:** Indicates the acceptance of a tunneled message with minimum encapsulation by the agent.
6. **G:** Indicates GRE encapsulation (Generic Routing Encapsulation).
7. **R:** (Reserved) field set as zero or ignored on reception.
8. **T:** Indicates that FA supports reverse tunneling.
9. **Reserved:** Not currently used fields, reserved for future purpose.
10. **FA Care of Addresses:** Addresses provided by the foreign agent.

The **type** field for prefix length extension is set at 19. **Length** is the value for the 'Num address:' field in the ICMP router advertisement part of the agent advertisement that indicates the number of router addresses advertised in the message. **Prefix length** defines the number of bits in the network prefix of the corresponding router addresses advertised in the ICMP router advertisement part.

12.7.5 Registration

When away from home, the mobile node registers (binds) it's care-of-address (CoA) with its home agent address (HoA), i.e., HoA CoA. The mobile node uses a special registration process to keep its home agent informed about its current location. Whenever a mobile node moves from its home network to a foreign network, or from one foreign network to another, i.e., the MS detects its movement to a new position, it chooses a foreign agent on the new network and uses it to forward a **registration message** to its home agent. The movement detection may be done in many ways—firstly, by viewing the remaining lifetime field in the agent advertisement message and secondly, by comparing network prefixes (network prefix of the old with the new IP subnet to which the mobile is moved to). Thirdly, the MH may detect the change of radio access points or the radio channel.

On entering a new network, the MH first acquires a temporary CoA from the visited network. Then the mobile needs to pass this information of CoA to the home agent. For MIPv4, the MH sends a **registration**

request with the CoA information. When the HA receives this request, it adds the necessary information to its routing table and checks the authenticity of the MH. If positive, the HA approves the request, and sends a **registration reply** message back to the MH (Fig. 12.18). These registration request and reply messages are sent over the UDP by using 434 ports.

An MH may register its current CoA with its HA directly or via the FA as depicted in Fig. 12.19. Registering through FA allows the visited network to enforce policies on network access and accounting.

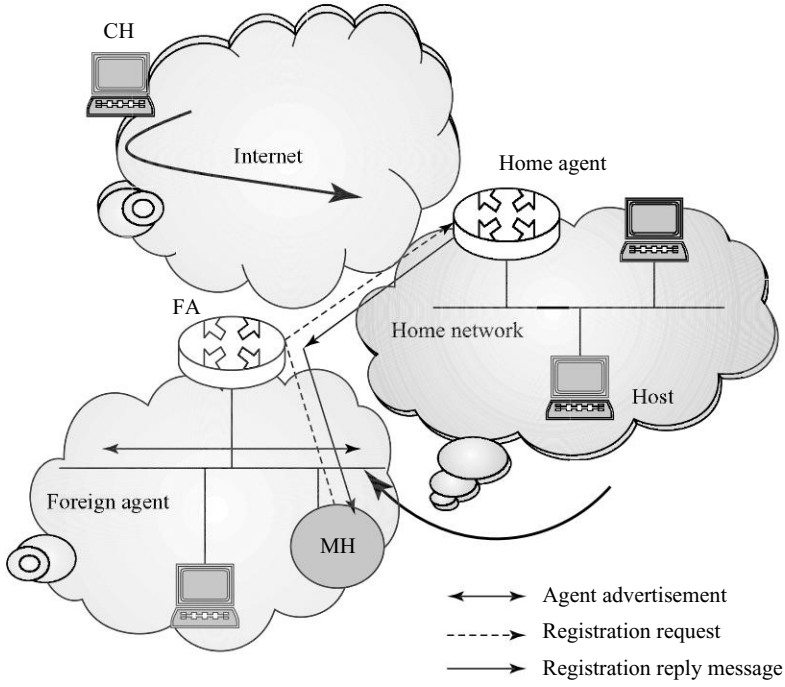


Fig. 12.18 *Mobile IPv4 registration process*

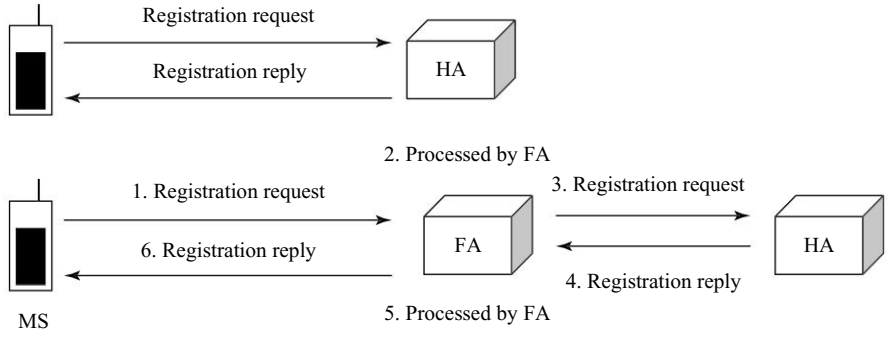


Fig. 12.19 *MIPv4 registration message flow*

12.7.6 MIPv4 Registration Request/Reply Message Format

All the registration requests are required to be authenticated by the home network, whereas all the registration reply messages are authenticated by the mobile node. This is required to protect against security attack. Any

mobile node can also send the registration request to discover the home agent, renew the registration due to expiration, deregister with the home agent when it returns to the home and also to discover the mobile node's home address if it is not configured with the HoA. The registration request and reply message formats for MIPv4 are given in Figs. 12.20 and 12.21 respectively.

The **type** field indicates whether the message is registration request or registration reply. The flags S, B, D, M, G, r, T, x are defined as follows.

1. **S:** Indicates simultaneous binding. If the S bit is set then MH may request for binding of multiple care-of-addresses (CoAs) to its home agent. Upon receiving packets by HA destined to MH, it will tunnel a copy of the packet to each CoA.
2. **B:** Stands for broadcast datagram. If B is set, the MH requests the HA to tunnel any broadcast datagram received on its home network.
3. **D:** Decapsulation b y MH. If the D bit is set, the MH will decapsulate the datagram sent to its CoA.
4. **M:** Minimal encapsulation. If set, the HA will use minimal encapsulation given in [9] for tunneling data to MH.
5. **G:** GRE encapsulation [8]. MH requests the HA to use GRE encapsulation.
6. **R:** Set to zero and ignored. Reserved bit.
7. **T:** Used for reverse tunneling.
8. **X:** Always set as zero and ignored on reception.

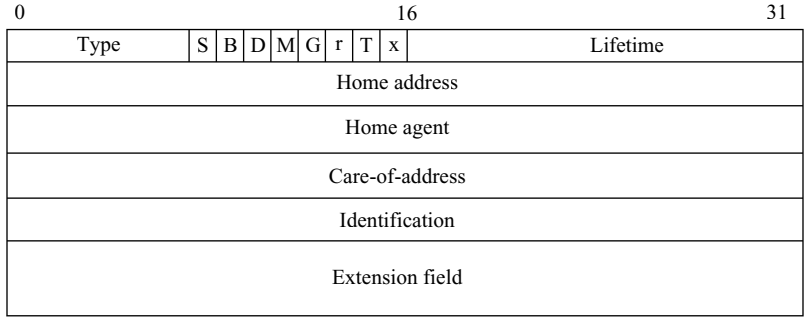


Fig. 12.20 *MIPv4 registration request packet format*

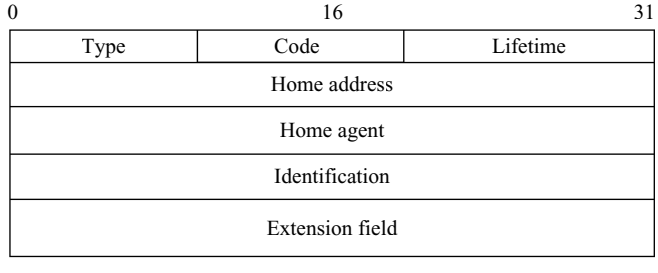


Fig. 12.21 *MIPv4 registration reply packet format*

Lifetime is the time in seconds for any registration. If the value for this field is zero, it means mobile node has been deregistered. If the registration request is not accepted then the lifetime field is ignored.

Home address is the address of the MH home address. The **home agent** is the IP address of the MH's home agent. The **identification** field contains a 64-bit number to check the match for registration request and reply. The value is based on the identification field in the registration request message received from the MH, and the way of reply protection used for security reason between the MH and it's HA.

When the MH is pre-configured with its home address then the home address field is the address of HoA in the Registration Request message; otherwise it is set as 0.0.0.0 for dynamic allocation of home address. The MH specifies the Network Access Identifier (NAI) to the request message using the mobile node NAI

extension. Upon reception of a registration request with this NAI extension, the HA can assign a valid home address to the MH that will be sent back in the registration reply message by the MH.

The **home agent field** is the IP address of the MH if the MH knows it. The MH can also use **dynamic home agent address resolution** for discovering it's HA, in case the MH does not know its HA. The MH uses the subnet-directed broadcast address of the MH's home network to the home agent field in the registration request message. Under the home network, all nodes will receive this request with broadcast destination address and in the reply message, will send the registration rejection to the MH that contains the address of home agent. In this way, the MH can learn about its home agent address.

The **care-of address field** contains the CoA of the MH. It changes at each new point of attachment and can be thought of as the MH's topologically significant address; it indicates the network number and thus identifies the mobile node's point of attachment with respect to the network topology. **Extension fields** are meant for future enhancement of MIPv4. These are optional fields carried by the registration request message.

In the registration reply message, the **code** field is the result generated by the corresponding registration request message. Other fields have the same meaning as in the registration request message.

12.7.7 Tunneling

Tunneling is the process of transferring datagrams by the home agent to the care-of address. In Mobile IP, the home agent redirects packets from the home network to the care-of address by constructing a new IP header that contains the MH's care-of address as the destination IP address. This new header then encapsulates the original packet, causing the MH home address to have no effect on the encapsulated packet's routing until it arrives at the care-of address. Such encapsulation is also called *tunneling*, which suggests that the packet burrows through the Internet, bypassing the usual effects of IP routing. The tunneling operation of a Mobile IP is shown in Fig. 12.22.

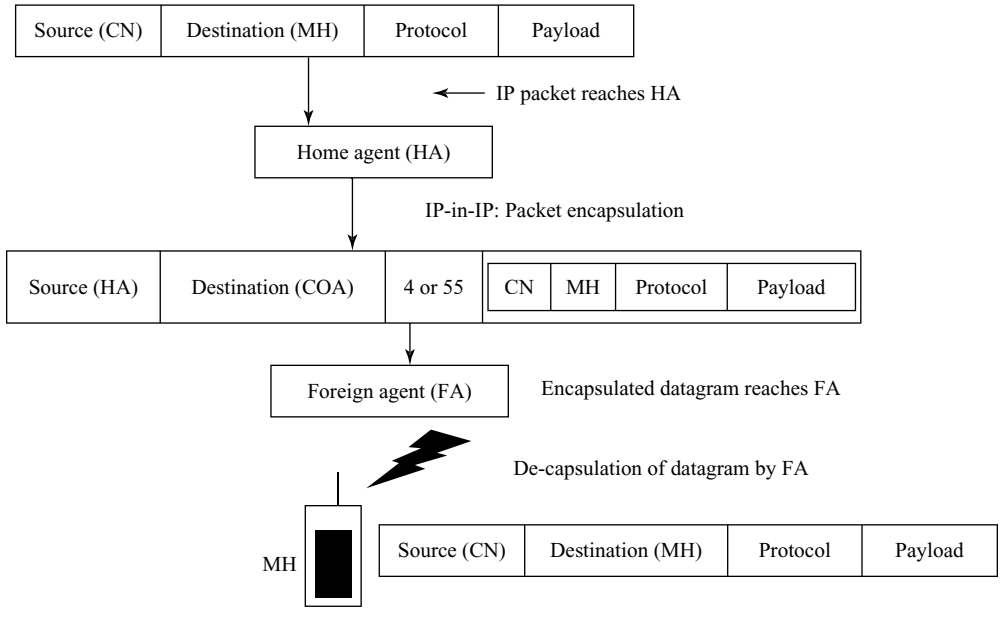


Fig. 12.22 *Tunneling mechanism in MIPv4*

The tunnel source is the home agent. It inserts a new IP header or tunnel header in front of the IP header of any datagram addressed to the MH's home address. The MH's CoA is used as the destination address. The tunnel source is the IP address for HA. The presence of 4 in the tunnel header indicates that the next higher-level header is an IP header. The encapsulated packet is intercepted by the FA, removes the tunnel header (de-capsulation) and delivers the packet to the MH. When the tunnel header uses minimum encapsulation, the protocol number used is 55. Using minimum encapsulation reduces header overhead though it needs added complexity.

12.7.8 MIPv4 Reverse Tunneling

A reverse tunnel is a tunnel that starts at the mobile node's care-of address and terminates at the home agent. Consider a situation when the mobile node sends packets from a visited network. But the source IP address in the packet is the mobile node's home address. The IP access router in a visited network may reject the packet whose IP address is not part of the visited network. This is generally known as *ingress filtering*. The packets sent by a mobile node may not be passed through the access routers of the visited network if ingress filtering is implemented. In this situation, reverse tunneling provides a solution by setting up a reverse tunnel from the mobile node's care-of address to the home agent so as to ensure a topologically correct source address for the IP data packet. A mobile node can request a **reverse tunnel** between its foreign agent and its home agent when the mobile node registers. Fig. 12.23 shows the Mobile IP topology that uses a reverse tunnel.

The mobile node requests the reverse tunneling service when it registers through the selected FA that supports reverse tunneling by setting a 'T' flag in the MIPv4 registration request. As a result, FA can establish a reverse tunnel to the mobile node's HA.

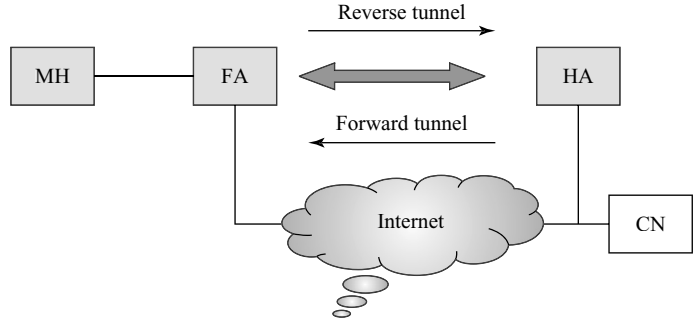


Fig. 12.23 *Reverse tunneling in MIPv4*

12.7.9 MIPv4 Triangular Routing

There are three steps for IP-packet transfer in MIPv4, generally known as triangular routing.

1. A datagram to the mobile node arrives on the home network via standard IP routing.
2. The datagram is intercepted by the home agent and is tunneled to the care-of address.
3. The datagram is de-tunneled and delivered to the mobile node.

For datagrams sent by the mobile node, standard IP routing delivers each to its destination. Figure 12.24 illustrates the triangular routing procedure.

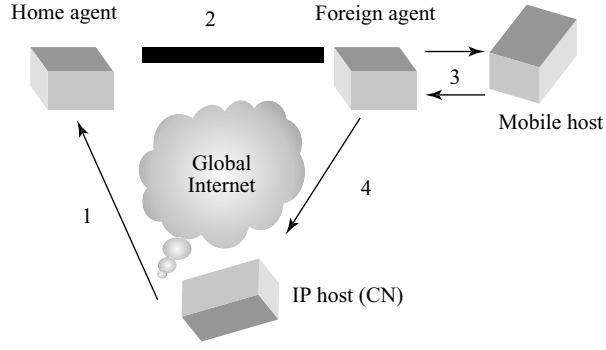


Fig. 12.24 *MIPv4 triangular routing*

12.8 PROBLEMS AND LIMITATIONS OF MIP

There are several problems related to Mobile IP operation. A few of the problems are discussed in this section.

1. Large signaling overhead due to large number of registration update
2. Triangular routing
3. Large handover latency
4. No scalability support
5. Inefficient use of existing public address space—since HA uses the mobile agents current CoA, we need at least one global address per subnet (for FA).

Increase Number of Registration Update Every time a mobile host moves beyond its limits of the link layer connectivity, a registration update is required for the host with its home network agent (HA).

It cannot support fast handoff and seamless mobility in handoff-intensive environments, such as when the MN moves within a small geographical area remotely located with respect to its home network (HN). This leads to large signaling overhead and suboptimal performance as all the traffic outside the home network has to go through the mobile node's home agent resulting in a performance bottleneck. When the foreign network is far away from the home network, the registration process takes a long time. As a result, the handoff time may also be high, which is unacceptable for real-time voice and multimedia communication.

Triangular Routing could Introduce Long End-to-end Delay and Produces Routing Inefficiencies This necessitates MIPv4 route optimization technique. The base Mobile IP specification has the effect of introducing a tunnel into the routing path followed by packets sent by the correspondent node to the mobile node. As the packets from the mobile node can go directly to the correspondent node with no tunneling, this asymmetry is caused by triangular routing. A single leg of the triangle goes from the mobile node to the correspondent node, and the home agent forms the third vertex controlling the path taken by data from the correspondent node to the mobile node. To alleviate triangular routing, techniques of route optimization are required, requiring changes in the correspondent nodes. In MIPv6 (Mobile IP version 6), the problem of triangular routing is handled by sending a binding update message to the corresponding node so that it can send data directly to the mobile in the visited network by passing the home agent.

Security is of great concern in any application running Mobile IP. Implementation also becomes tough as firewalls block all classes of incoming packets that do not meet specified criteria or permission. Many border routers may also discard packets coming from any correspondent node to the enterprise network if the packets do not contain the source IP address configured with that enterprise's internal networks. Generally, mobile nodes use their home address as the source IP address that creates difficulty. Using a home agent with reverse tunneling in Mobile IP may solve the problem but generates the problem of triangular routing.

Scalability A home agent may have performance bottleneck because of traffic overflow. All user traffic outside the home network will have to pass the mobile nodes' home agent that makes the HA traffic a bottleneck. As the number of mobile nodes increase, this effect will be more prominent and scalability will be a concern.

12.8.1 MIPv4 Route Optimization

The route optimization technique in MIPv4 is implemented to remove the problem of triangular routing. Fig. 12.25 shows the route-optimization technique. It enables the correspondent node to address packets to a mobile's current CoA directly so that involvement of the home agent is not required.

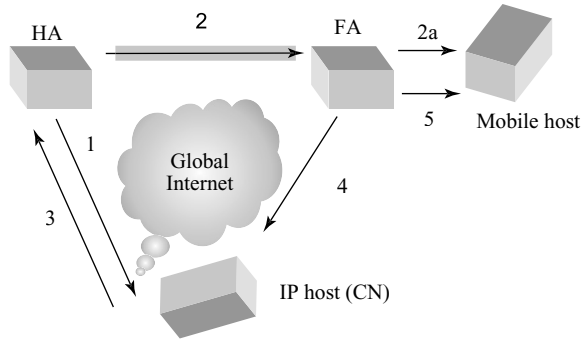


Fig. 12.25 *MIPv4 route optimisation*

Steps for MIPv4 route Optimization

1. Initial packets send to mobile's home address
2. Packets tunneled to mobile's CoA
3. Packets delivered to mobile
4. Binding update with CN
5. Subsequent packets tunneled to mobile's CoA directly
6. Packets delivered to mobile

The correspondent node (CN) is to be aware of the mobile node's current CoA and then packets are directly tunneled to its CoA. The CN maintains a binding cache that maps the home addresses of the mobile node with their CoAs. At the time of packet delivery, the CN first searches the binding cache for the CoA. If found, the packets are tunneled to the CoA directly, otherwise, the packets are sent to the home agent address. The home agent will send the CN, the CoA of a mobile by sending the binding update message. The binding update message contains the home address and CoA of the mobile. A security association is required between the CN and HA to accept binding updates. But maintaining security association is a critical limitation of MIPv4 for route optimisation in large networks. Thus, scalability will be the great concern.

Summary

In this chapter, the basics of Internet protocol and Mobile IP have been discussed which are the building blocks of mobility management and packet forwarding across networks. The brief overview of Mobile IP shows the potential market in mobility management for wireless networks, though it has some limitations. Security is one of the major concerns and needs active research attention. The problem of higher location update, signaling overhead, high handoff latency are partly alleviated by using hierarchical network architecture. Mobile IP version 6 (MIPv6) will become the mobility management platform with some additional features like in-built route optimization, security, ingress filtering, etc. MIPv6 and hierarchical Mobile IP architecture will be discussed in a separate chapter along with mobility management issues for wireless networks.

References

[1] Forouzan, B., *Data Communications and Networking*, McGraw-Hill, Special Indian Edition, Third Edition, 2004.

[2] Bannister, J., P. Mather, and S. Coope, *Convergence Technologies for 3G Networks*, *IP, UMTS, EGPRS and ATM*, John Wiley, 2004.

[3] Cheng, Chen J. and T. Zhang, *IP-Based Next Generation Wireless Networks*, Wiley, 2004.
[4] Perkins, C., *IP Mobility Support*, ed., IETF RFC 2002, Oct. 1999.
[5] Perkins, C., *Mobile IP: Design Principles and Practice*, Addison–Wesley Longman, Reading, Mass., 1998.
[6] Perkins, C., *Mobile IP*, IEEE Comm., Vol. 35, No. 5, pp. 84–99, 1997.
[7] Deering, S.E., *ICMP Router Discovery Messages*, ed., IETF RFC 1256, Sept. 1991.
[8] Perkins, C., *IP Encapsulation Within IP*, IETF RFC 2003, May 1999.
[9] Perkins, C., *Minimal Encapsulation Within I, IETF RFC 2004, May 1999.*
[10] Deering, S., and R Hinden, *Internet Protocol Version 6 (IPv6) Specification,* IETF RFC 2460, Dec. 1998.
[11] Johnson, D., and C. Perkins, *Mobility Support in IPv6,* IETF draft, draft-ietf-mobileip-ipv6-15.txt, July 2001.
[12] Campbell, A. T. *et al*., *Comparison of IP Micromobility Protocols,* IEEE Wireless Commun., pp. 72–82, Feb. 2002.
[13] Salkintzis, *Mobile Internet–Enabling Technologies and Services,* CRC press, 2004.
[14] Sollman, H. *et al*., *Hierarchical MIPv6 Mobility Management,* IETF draft, draft-ietf-mobileip-hmipv6-05.txt, July 2001.

Questions for Self-Test

12.1 IP-based protocols are independent of the underlying technology.
a. True b. False

12.2 Mobile IP supports seamless mobility.
a. True b. False

12.3 MIP supports transparency above IP layers.
a. False b. True

12.4 MIP maintains active TCP connections and UDP port bindings.
a. False b. True

12.5 Indicate which of the following are applicable for IP services.
a Connectionless b. Best effort
c. Fragmentation d. Reassembly

12.6 TOS field in IPv4 defines different QoS parameters.
a. True b. False

12.7 Identification field is required for re-assembling the fragmentation datagram.
a. False b. True

12.8 TOL field prevents misrouted data for looping option header allowing security handling.
a. False b. True

12.9 IPv6 header is 60 bytes long.
a. False b. True

12.10 A flow label can support special handling of IPv6 routers.
a. False b. True

12.11 An IP holds a unique and universal address.
a. True b. False

12.12 Classes A, B support multi-casting.
a. True b. False

12.13 Class D is for unicast routing.
a. True b. False

12.14 What are the different addressing schemes for IPv4?
12.15 Class A is designed for large organizations with a large number of hosts; Class B is for medium-size organizations and class C is for small organizations. Why?
12.16 Subnetting is used to handle efficient network address. How?
12.17 Subnet masking is required to get a network number. Explain how?
12.18 Default masking is a _____ bit number.
12.19 Next header option in IPV6 connects _____ protocols.
12.20 TCP is a _____ to _____ protocols.
12.21 Which of the following are suitable for TCP?
a. Connection oriented
b. Reliable
c. Layer 4
d. Congestion control
12.22 How many header fields are there in IPv4? Discuss briefly the functions of each field.
12.23 Why is port number needed in TCP header? Why is sequence number needed?
12.24 How is packet flow maintained in a TCP connection?
12.25 Which are true for UDP connection?
Connectionless/reliable/simpler than TCP/supports faster transmission
12.26 UDP is suitable for _____ and _____.
12.27 BGP is a routing protocol for _____.
12.28 RIP, OSPF is _____ protocol.
12.29 An IPV6 datagram is composed of a base header and a _____.
12.30 _____ header add functionally to the IPv6 datagram.
12.31 Discuss the address schemes in IPV4 and IPV6 protocols.
12.32 Find the netid and hostid of the following IP addresses:
192.10.25.30; 127.57.14.3; 117.20.57.3
12.33 Determine the IP class:
236:15:5.6
129:100.50.5
111.36.2.8
12.34 What is the purpose of including the IPv4 header and the first 8 bytes of datagram data in ICMPv4 messages?
12.35 What is the size of ARP packet in IPv4 header? Why is IPV6 gaining importance over IPv4?
12.36 Discuss the IPv6 addressing schemes.
12.37 IPv6 header is much simpler than IPv4 and of standard 40-bytes length. Is this statement true? What are the greater advantages over IPv4 obtained?
12.38 What is the role of Address Resolution Protocol?
12.39 What is IP subnetting? Why is it required?
12.40 What is called subnet masking?
12.41 What is NAT? How does NAT help in address depletion?
12.42 Explain the addressing scheme of Domain Name System (DNS)?
12.43 What is fully qualified domain name?
12.44 What are the main functional entities for Mobile IP?
12.45 Describe the Mobile IP operation with diagram.
12.46 How is the problem of triangular routing handled in mobile IPv4?
12.47 What are the limitations of MIPv4?
12.48 How is data transferred to a mobile node in MIP operation?

Mobility Management Issues: Role of IP on Wireless Networks

13

Introduction

The next-generation wireless network will be based on Internet protocols. A wireless IP network uses IP to support key aspects of network operations like network layer routing, transport of user packets, mobility management at the network layer or higher layers, and voice and multimedia services. The aim is to get all-IP wireless networks that would support IP at the network layer or above in the core network or in the radio access networks. This would create globally accessible Internet services for wireless paradigms. The IP-based protocols are independent of underlying radio technologies and support seamless mobility across different radio networks for global roaming. All computers in the Internet world are able to understand IP.

The convergence of fixed, mobile and data networks offer new possibilities to network operators in making efficient use of the network resources for end users, and in turn reap profits by using new applications and a transparent billing system. This could be possible only when a single network architecture is administered. Making the heterogeneous platform available to all IP-based networks allows the development of new applications which make use of the interaction between voice calls and Internet connections at the same time within the same application.

GSM is widely used all over the world, but GSM transmission is purely circuit switched and not resource efficient. The data rate is very low compared to that of fixed networks. Only one time slot for uplink and one time slot for downlink are available for each user. GPRS is a more efficient network platform for data transmission as it is based on packet-switched connections. Thus, a new packet-oriented core network consisting of SGSN and GGSN on an IP network is required. GGSN is the interface directly connected to the external packet data network, i.e., intranet, Internet and PDN. But the data rate in GPRS is not enough for the present volume of multimedia communications. For third generation UMTS networks a larger frequency band is used for the radio interface to obtain a higher transmission rate. Use of code division multiple accesses for the UMTS network provides higher flexibility for the distribution of radio resources over the radio interface.

As described in Chapter 11, UMTS R99 uses the IP for data transport within the core network. The GPRS packets are routed between the user equipment and external packet networks. But GPRS tunneling protocol (GTP) is used across the UMTS core network. In UMTS R99, the core network is divided into two parts, the circuit-switched networks (all MSC/VLR and GMSC) and the packet-switched network (all SGN/SLR and GGSN). The network operators need to install, configure and maintain two different types of core networks. The idea for UMTS R4 is to merge these two networks into a single IP-based core. In UMTS R4, IP is used to transport both user and control data across the UTRAN through the core network and beyond. Voice traffic within the core of the R4 network passes using voiceover IP or voiceover ATM (asynchronous transfer mode) technology. But in R5 or R6, the UMTS network is evolved to accomodate to all IP architecture. So both the UTRAN and core network can use IP transport.

One of the salient features of wireless communications is the flexibility to support user roaming. Roaming causes two phenomena–handoff and location update. To maintain connection continuity, ongoing sessions are to be transferred at the new access points of the roaming user. Again, the new location of the mobile has

to be tracked by the home location register of the network where the mobile's permanent address resides in order to deliver messages addressed to the mobile. So, efficient mobility management is required that consists of two parts–hand-off management and location management. The wireless/cellular industry is considering Mobile IP as one technique to implement IP mobility for wireless data. The working group will endeavor to gain an understanding of data service in cellular systems such as GPRS, UMTS, CDMA2000, and interact with other standard bodies that are trying to adopt and deploy Mobile IP Wireless Gateway protocols in these contexts.

13.1 IP FOR GPRS AND UMTS R99

GPRS is the packet-switched service for UMTS and GPRS networks. The word GPRS is often used for a GSM network that supports GPRS packet data transportation. The radios for GSM and UMTS are different, but the IP network looks the same.

To carry data between the UE and the hosts on the Internet or other external PDN for UMTS/GPRS networks, mobile users use IP as a transport mechanism. The second option is to use the GPRS tunneling protocol (GTP). Using GTP ensures security from hackers operating from both the mobile and fixed network end.

User and transport planes are completely independent, i.e., the transport plane can run on a different IP version than the user plane. UTRAN and core network transport can theoretically run on different IP versions as shown in Fig. 13.1.

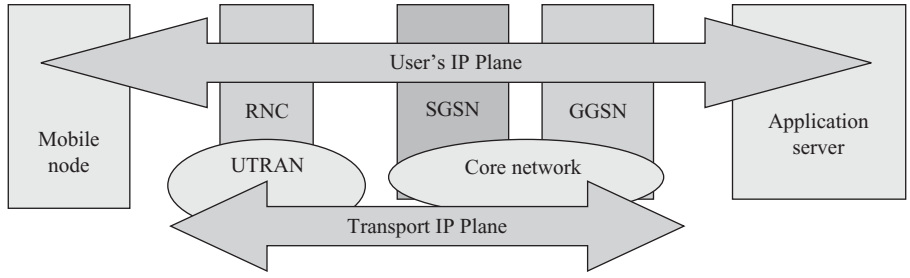


Fig. 13.1 *IP for UTRAN and Core Network in UMTS*

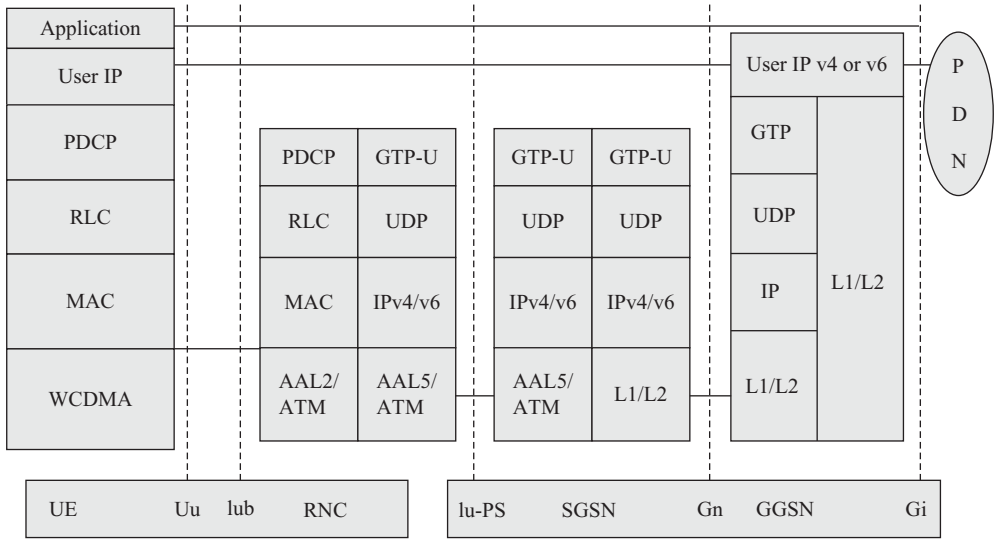


Fig. 13.2 *Protocol stack for user plane in UMTS R99 between the mobile and GGSN*

Figure 13.2 shows the protocol stack in the user plane between the mobile and GGSN in UMTS R99. The role of IP is limited to packet-switched domain for R99. The UMTS Radio Access Network is based on ATM. It uses the ATM Adaptation Layer 2 (AAL2) for user traffic and AAL5 for carrying signaling traffic. The IP is used as transport (GTP) in the PS domain between the SGSN and GGSN and over the Iu interface between the RNC and SGSN.

PDCP The Packet Data Convergence Protocol is used to transport higher-layer packets. It supports Point-to-Point (PPP), IPv4 and IPv6 as the higher-layer protocols. Header compression is the primary function of PDCP because IP protocol introduces large header overheads.

13.2 PROTOCOL REFERENCE MODEL FOR UMTS PS DOMAIN

The protocol reference model for the PS domain is shown in Fig. 13.3. To understand the role of IP on the GPRS/UMTS networks, we need to understand the basics of interfaces and their roles between the nodes on the PS domain. The main interfaces for supporting packet-switched services are the Gi, Gn, Gp, Gc, Gr interfaces inside the packet core network, whereas Iu interfaces connect the RAN on the PS core network.

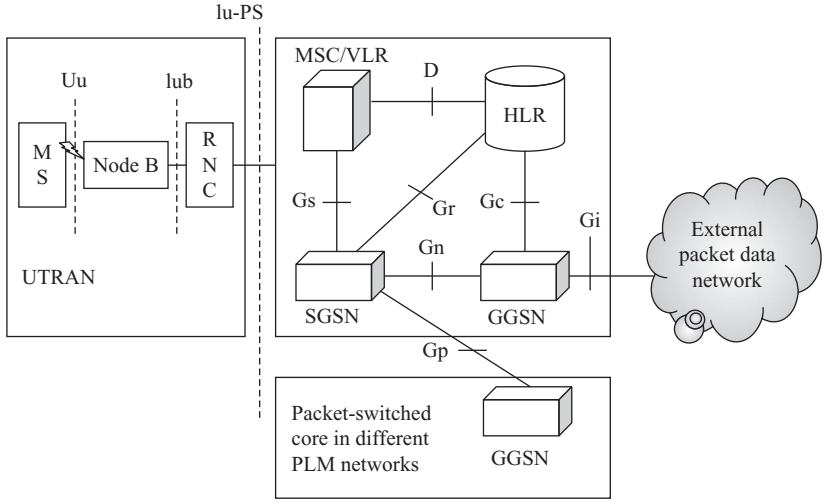


Fig. 13.3 *Protocol reference model for the PS domain in UMTS network*

13.2.1 Packet-Switched (PS) Domain Protocol Stacks: Role of Interfaces

Interface Gi is a standard IP interface between a GGSN and the IMS and other IP networks. The Gi interface uses IP as the network-layer routing protocol. To an external network, GGSN acts like a regular IP router and the GPRS/UMTS PS domain acts as an IP network. A mobile uses Packet Data Protocol (PDP) to exchange user packets over the PS core network domain with other mobiles within the same network or in other IP networks. As standard IP routing protocol is used over the Gi interface, the data and control planes are identical. The Gi interface protocol stack is shown in Fig. 13.4.

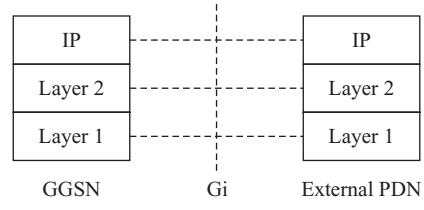


Fig. 13.4 *Gi interface protocol stack for GPRS/UMTS network*

The Gn interface is used between the SGSN and GGSN as well as between SGSNs in the same PLMN (Public Land Mobile Networks), whereas the Gp interface is used between the SGSN and a GGSN in a different PLMN. The protocol for both the control plane (signaling) and user plane (user data packets) over Gn and Gp interfaces is the GPRS tunneling protocol GTP. GTP is also the tunneling protocol between the RAN and SGSN in the core network. Figure 13.5 shows the UMTS Gn and Gp interface protocol stacks.

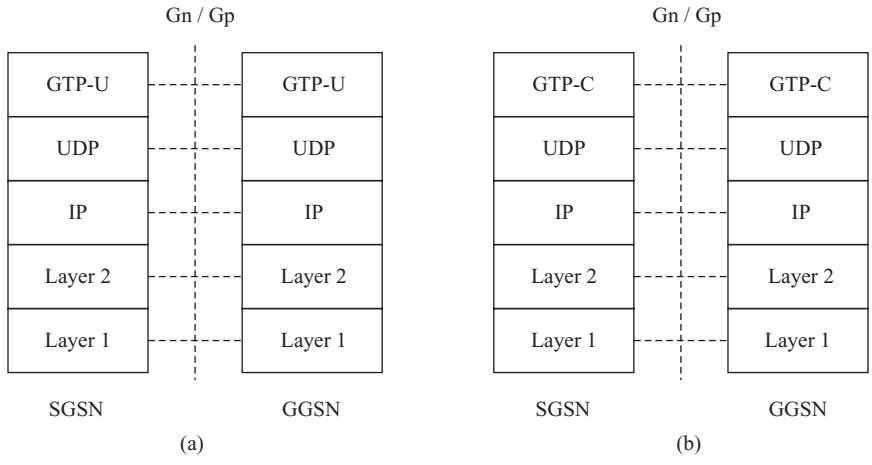


Fig. 13.5 *UMTS Gn and Gp interface protocol stack for (a) user plane, and (b) control plane*

13.2.2 GTP Tunnel

It is to be mentioned here that GTP is not only a tunneling protocol, but also a signaling protocol used to support PDP context activation, deactivation and modification are an essential part of the GPRS/UMTS networks. Inside the PS core network, GTP manages mobility related functionality such as host specific routes for the mobile while moving. The two sets of messages and procedures are GTP-U (GTP User Plane) and GTP-C(GTP-Control Plane). GTP-U is responsible for managing PDP contexts, location management and mobility management, whereas, GTP-C manages control tunnels, their creation, modification and release. For every active PDP context, one GTP-U tunnel is established between the SGSN and GGSN.

Thus, GTP tunnel management messages help to activate , modify and release PDP contexts and their associated GTP tunnels between an SGSN and a GGSN. GTP-location management messages (For example, GTP-C) are used to retrieve location information from the HLR by the GGSN. But HLR is designed mostly in a circuit-switched manner. The Mobile Application Part (MAP) protocol is generally used to exchange information from the HLR to any other network nodes. So, for location management, GTP messages need to be converted to MAP messages with the help of a protocol converter. Figure 13.6 illustrates the protocol stack for supporting this scenario.

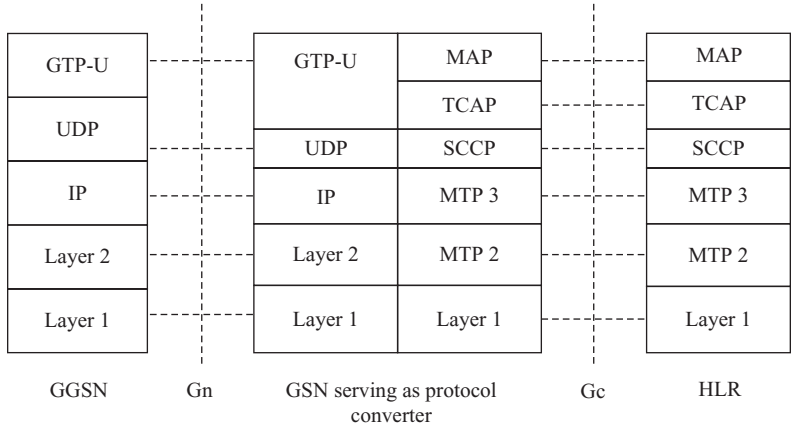


Fig. 13.6 *UMTS protocol stack on GTP between the GGSN and HLR*

GTP Mobility Management Messages (GMM) are used between the SGSNs within the network. Mobility-related messages are transferred from one SGSN to another while the mobile node performs the GPRS attach procedure, routing area and location update procedures during the handoff occuring from one SGSN to another. GTP is an essential protocol of the GPRS/UMTS network. There are several protocols proposed by IETF that tunnels IP packets over any non-IP or IP network. Mobile IP is one such alternative protocol that uses IP-in-IP tunneling. For implementing Mobile IP on the GPRS networks, it is required that the GGSN always knows the serving SGSN of the mobile node. So, it plays an important role in mobility management.

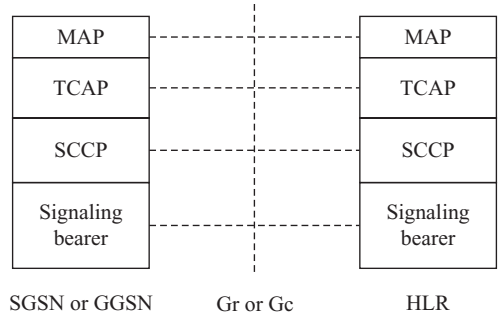


Fig. 13.7 *Protocol stack between SGSN or GGSN and HLR*

The Gr interface between the SGSN and HLR and the Gc interface between the GGSN and HLR use an identical protocol stack that uses MAP over Gr and Gc for signaling. Again, MAP is implemented over the SS7 Transaction Capabiliities Application Part (TCAP). TCAP is transported over SS7 SCCP that is implemented on ATM connections for UMTS networks.

13.2.3 Iu-PS Interface and Mobility Management

The Iu–PS interface provides procedures for supporting handoff between the RNCs of the UTRAN. One of the main functionalities that Iu-PS provides is the serving RNS relocation that is needed to be transferred from the serving RNS side of a Radio Access Network Application Part (RANAP) connection from one RNS to another during handover procedure. RANAP runs over the SS7 Signaling Connection Control Part (SCCP). SCCP is a transport-layer protocol similar to UDP and TCP over an IP network. The difference is that SCCP runs over ATM using ATM Adaptation Layer 5 (AAL5). To transport RANAP messages for each individual mobile, a separate SCCP connection is needed. Figure 13.8 illustrates the user and control plane protocol stacks.

User packets are transported over the Iu–PS interface using GTP-U tunnels that are implemented over UDP/IP. Again, the UDP/IP stack for transporting GTP-U packets can be implemented over any lower level technology.

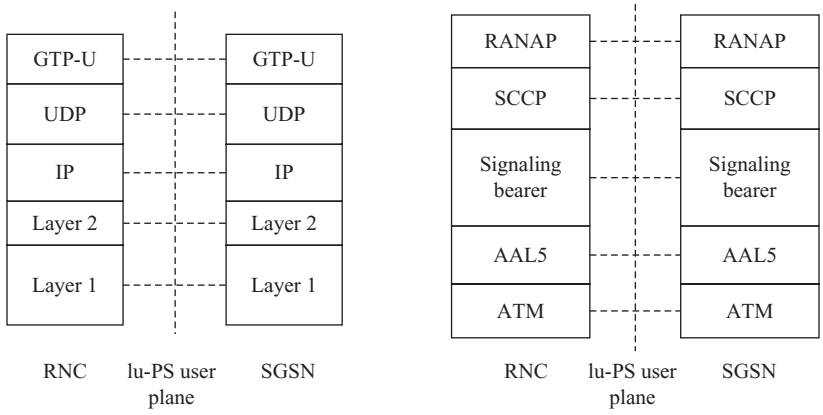


Fig. 13.8 *Iu-PS domain interfaces for UMTS network*

13.3 PACKET ROUTING AND TRANSPORT OF USER DATA IN UMTS NETWORKS

IP is the main transport mechanism inside the packet-switched core network between the nodes RNC to SGSN, SGSN to GGSN, GGSN to GGSN and also between SGSNs. Though IP routing is used to transport

data between the GGSNs, but the GPRS specific tunneling protocol, GTP, is used in between SGSN and GGSN and also between RNC and SGSN.

User data packets from any source mobile node are first sent to GGSN that finally forwards the packets to its final destination. Even when two mobiles are at the same GGSN, packets first come to GGSN. GGSN plays the central role for routing all the user packets.

Tunneling means encapsulation of packets—putting one packet into another. Based on the destination that is kept in the header information, the encapsulated packets are routed. GTP is the main tunneling protocol in GPRS/UMTS networks. User packets are tunneled between the RNC and SGSN, SGSN and GGSN and GGSN to GGSN. GTP enables GPRS specific protocols instead of IP routing and IP mobility management protocols inside the PS domain. Tunneling makes the transport of the user packets accessible via protocol independence in the external packet data networks.

Packet transport between the mobile node and GGSN following the host-specific information depends on the specific PDP context created for an individual mobile. GGSN maintains the routing information of the serving SGSN for an active PDP context. When the mobile node roams between the SGSNs, GPRS mobility management (GMM) is used to maintain the routing from SGSN to GGSN, updating the routing area. Thus, host-specific routing does not follow IP routing, but uses prefix based routing. Figure 13.9 depicts the user packets routing in the PS domain.

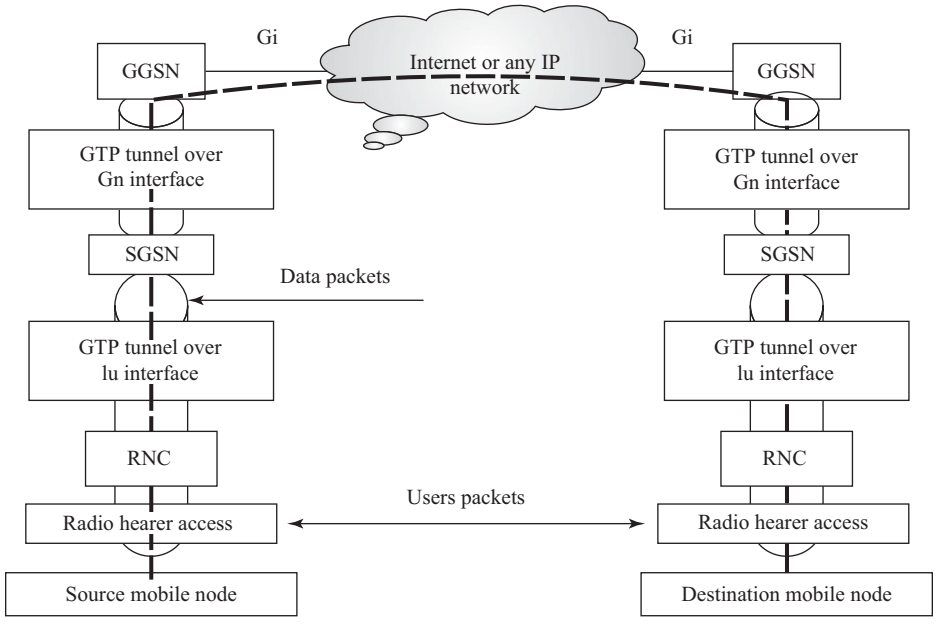


Fig. 13.9 *Packet routing through PS domain*

13.3.1 Roaming In GPRS Networks

While roaming in GPRS home, GGSN uses GPRS Roaming Exchange (GRX) to transfer data traffic between the operators. It allows access in home services. GRX is not connected to the Internet for security reasons. Virtual Private Network (VPN) is used between the GPRS operators, designed to carry inter-operator traffic, and network elements sit in their own network. Intra-operator networks are usually connected via an inter-operator network GRX. Due to GPRS security model, the GPRS intra and inter-operator networks are not connected to the Internet. In an interconnected GPRS network, different GPRS elements should have the capability to connect directly to each other. For that, network element addresses have to be unique and can use public IP addresses.

Figure 13.10 illustrates the roaming scenario in GPRS networks.

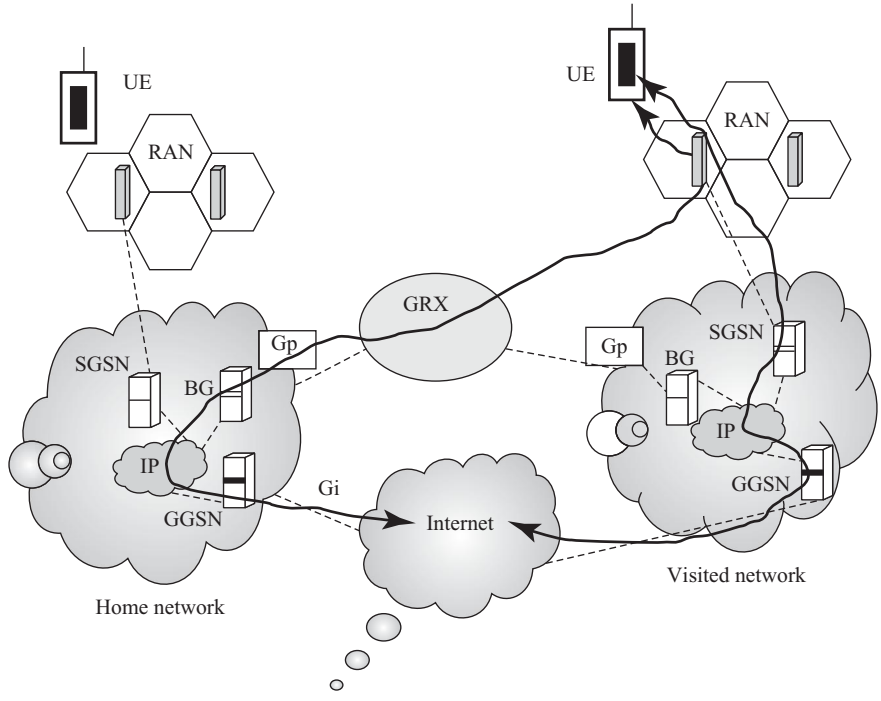


Fig. 13.10 *Roaming scenario in GPRS network*

13.3.2 Configuring PDP Addresses on Mobile Stations

Activation of PDP (Packet Data Protocol) means acquiring an IP address to send or receive packet data by the mobile user over the GPRS/UMTS network. When the User Equipment (UE) attaches to the network, the SGSN creates a mobility management state containing information pertaining to mobility and security for the UE.

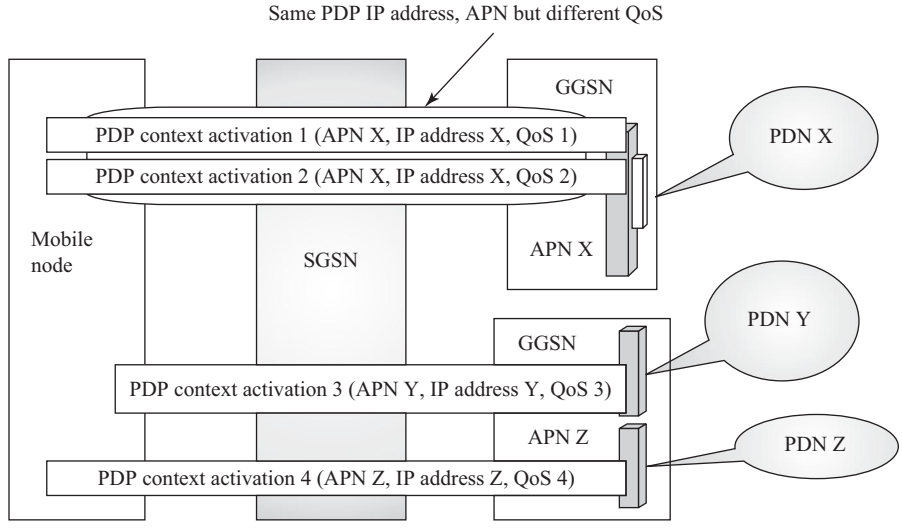


Fig. 13.11 *PDP context with external PDN*

At PDP Context Activation, the SGSN and GGSN create a PDP context, containing information about the packet data session (e.g., IP address, QoS, routing information, etc.). Each subscriber may activate several PDP contexts towards the same or different GGSNs. When activated towards the same GGSN, they can use the same or different IP addresses. Figure 13.11 illustrates the PDP context activation with the external PDN.

Access Point Name (APN) The APN is a logical name referring to a GGSN. The APN also identifies an external network. The syntax of the APN corresponds to a fully qualified DNS name. At PDP context activation, the SGSN performs a DNS query to find out the GGSN(s) serving the APN requested by the mobile terminal.

There are four possibilities of getting a PDP address (IP address).

1. The visited network may assign static IP address to which the mobile is currently attached. The visited network ensures that the assigned static PDP address is always routed to the mobile node's serving GGSN, which will forward the packets to the destined mobile. During the PDP context-activation process, the mobile node informs the static PDP address to the visited network to be used by the SGSN.
2. The visited network may dynamically assign an IP address to which the mobile is currently attached. The mobile serving GGSN in the visited network allocates the PDP address from the pool of IP address space during the PDP context-activation process.
3. A static PDP address is provided by an external IP network outside the visited PS domain. It may be any Internet Service Provider (ISP), intranet or the mobile node's home IP network. The assigned PDP address to the mobile node may not be part of the visited GPRS network. The external network and the visited GPRS network collectively ensure that the static PDP address correctly routes the packets to the serving GGSN.
4. A dynamically assigned PDP address is provided by an external IP network outside the visited PS domain. The mobile node informs the visited network about it's request to get a dynamically assigned IP address from the external PDN during the PDP context activation. The visited PS domain creates a PDP context without assigning a PDP address to the mobile. The mobile then sends requests to the external network to aquire a PDP address and receives a reply from the external PDN. The details of aquiring a PDP address dynamically is given in Chapter 9. The serving GGSN of the mobile in the visited network has to know the assigned PDP address for forwarding packets.

13.4 MOBILITY MANAGEMENT IN WIRELESS NETWORKS

Mobility management in wireless networks is an important issue. The role of Mobile IP in handling mobility, both in cellular and IP-based wireless networks, will be discussed in this section. The basics of mobility-management issues will be discussed first. Mobility management is a two-step process—Hand-off management and Location management. A wireless network has the flexibility to support user roaming, but the geographical coverage of a wireless network is limited. A backbone network, such as wireline or satellite network, is needed to extend the coverage. The user terminals can be mobile or fixed. With the development of mobile communications and Internet technology, there is a strong need to provide connectivity for roaming devices to continuously communicate with other devices on the Internet, at any time and at any place. The key issue of this vision is the know-how to support mobility in TCP/IP networks.

There is no single perfect solution so far; mobility support may require Internet architecture to have some general design considerations for any Internet2 mobility support solution. Internet mobility support refers to keeping ongoing-communication continuity when an IP-based device moves, i.e., changes its topological point of attachment to different networks. In order to provide such support, a number of fundamental issues arise, which can be summarized as the following requirements for Internet mobility support.

13.4.1 Mobility Classification

Mobility can be of different types:

1. **Terminal mobility:** It is the ability of the user to access the network even if the terminal is moved.
2. **User mobility:** It is the ability of a user to continue access to a network service under the same user identity when the user moves. That means a unique user identity is used at different terminals to access the network services.
3. **Service mobility:** Service mobility is the ability of a user to access the same services regardless of where the user is presently attached.

When a user terminal moves to a new location and connects to the network to get access at the new position, it is known as portability or discrete mobility. On the other hand, if a user moves and gets network access continuously, it is called continuous mobility.

13.4.2 Seamless Terminal Mobility Management

Users should be able to roam freely and seamlessly across the geographic boundaries of different constituent networks maintaining high data rate, best possible QoS and satisfactory connectivity with the application servers. Multimode terminals capable of automatically tracking and selecting the appropriate available underlying network play an important role in this context. Ample research activities in this field have taken place with prime concerns about reducing packet loss, signaling overhead and handover latency apart from increasing the throughput. IPv6, providing link-layer independent mobility management solutions, is unanimously accepted as the prime backbone of the future generation core network. Further protocol advancements in this context have implemented Mobile IPv6 (MIPv6) [3] and Hierarchical Mobile IPV6 (HMIPv6) [6]. The effect of other protocols like Location Independent Network Architecture for IPv6 (LIN6) [9,11], Transmission Control Protocol (TCP), User Datagram Protocol (UDP), Stream Control Transmission Protocol (SCTP), Datagram Congestion Control Protocol (DCCP), Multiple Address Service for Transport (MAST), Session Initiation Protocol (SIP) and Dynamic Domain Name System (DDNS) at the different layers are also being studied [5] in depth. A slow increase in the number of effective Personal Area Networks (PANs) has urged more research activities in tackling issues related to mobility management of roaming networks.

Location Management The process of tracking and maintaining the exact whereabouts of wireless terminals by the underlying system for possible connections when they are powered-on, powered-off or on the move is location management, which is a two-phase technique. Location tracking deals with tracking the exact location of the terminals, and location information storage [9] that maintains other location information like QoS capabilities, authentication and traffic. Roaming through a diversified range of networks has given rise to the need for efficient and integrated location management schemes.

Handover Management Efficient hand-off or handover management is another primary area of concern for maintaining global mobility. The most important function needed to support mobility is maintaining an ongoing communication while a mobile node (MN) moves and changes its point of attachment to the Internet. For seamless communication, a core technology, called *handover management*, is required. The main objective of handover management is to minimise service disruption during handover. While roaming, smooth handover of the mobile terminals is necessary for seamlessly maintaining an ongoing communication. Advancements in the various protocols designed play a prime role in this context.

Though MIPv6 was designed as the future standard mobility protocol for IP-based networks, it suffers from drawbacks like high packet loss, increased system load, high handover latency [13] and signaling scalability mainly due to the absence of location management hierarchy, absence of paging support and wastage of the mobile node's battery power. HMIPv6 promoted by the IETF as a further advancement to MIPv6 has till now proved to be a promising technology for the next-generation wireless networks by efficiently tackling the MIPv6 drawbacks.

Multi-homing under Heterogeneous Environment With the evolution of a wide range of wireless access techniques such as GPRS, WCDMA/UMTS, IEEE 802.11x, etc., to provide access to the Internet, the future mobile environment will be characterised by heterogeneous access networks, and the MN will be equipped with multiple interfaces supporting different wireless techniques. Thus, it is necessary to require multi-homing support by which the MN can access the Internet through multiple links simultaneously and select and switch dynamic links while moving.

Application Domains Internet mobility should also support current services and applications. That is to say, the mobility management mechanism should be transparent, without requiring changes to current services and applications.

Security Aspect Any mobility solution must protect itself against misuses of the mobility features and mechanisms. Snooping into unauthorised domain, stealing of legitimate addresses or flooding a node with a large amount of unwanted traffic violates the security. Therefore, security is an important concern when providing Internet mobility support.

In addition, there are performance requirements for mobile environments. While developing an Internet mobility solution, the performance metrics also deserve special attention. These are the following:

Handover Latency Time required from the last packet received via the old network to the arrival of the first packet along the new network during a handover.

Packet Loss The number of packets lost while maintaining communication during a handover.

Signaling Overhead The number of messages for the handover and location procedures.

Throughput The amount of data transmitted over a mobile Internet in a given period of time.

13.4.3 Basics of Handover Management

When a mobile node changes its point of attachment to the network, it moves from one network to another new network. This process is known as *handoff* or *handover*. During this process, the MN usually disconnects from the old network before connecting to the new network (especially if using a single interface) and thus there is a time when the MN has lost connectivity to the Internet. During this period, it cannot send or receive packets to the detriment of existing application sessions. The mobile can also change the radio channel from one base station to another within the same network administrative domain. It can also change the radio frequency channel from under the same base station during the signal quality degradation.

In a broad sense, hand-off may be hard or soft depending on how the mobile node gets data from the access network during the hand-off operation. Hard handoff occurs based on the two principles—make-before-break and break-before-make. The first one will occur when the mobile node sets up new connectivity with the target base station (access point) and then breaks the old connection. The second category describes the hand-off procedure when the mobile node tears down its connectivity to the old access point and then makes a connection to the new base station. It is to be noted that during hard handoff only one base station involvement is needed. Some data may be lost during the handoff.

For soft handoff, more than one access-point involvement is required. Different base stations send separate copies of the same data to the mobile simultaneously. Using signal-processing techniques, the mobile selects the correct data from multiple base stations. Similarly, the mobile may send several data copies from the same set to different base stations. The network then processes the data to get the correct single copy to send to the destination.

Handoff in IP-based wireless networks may work on different protocols at different layers such as the following:

1. **Physical layer handoff:** The mobile node changes its point of attachment at the physical layer itself like the change of a radio channel from one base station to other.

2. **Logical link layer handoff:** The mobile exchanges user IP packets over the logical link layer within the wireless network administrative domain.
3. **IP layer handoff:** When handoff occurs by changing network IP addresses. IP mobility does not depend on the lower layer protocols.

So, it can be mentioned that mobility at different protocol layers can be managed by different protocols. IP protocol layer handoff is divided into three parts. The entire geographical region is divided into domains, and domains are divided into subnets.

(a) *Intra-subnet handoff:* Handoff occurs within the same subnet. The mobile node does not change its IP address (one base station to other).

(b) *Inter-subnet or intra-domain handoff:* Handoff occurs in between the subnets, and change of IP address occurs due to a change of subnet address.

(c) *Inter-domain handoff:* Handoff occurs between two domains. The mobile node moves to the new network access having a different network IP address. This is also called global handoff. This type of handoff invokes time-consuming procedures like authentication and authorization.

Handoff Strategies In a wireless cellular network, while mobile node moves from one base station to another (physical layer handoff), its connection must be transferred from the current base station to the new target base station. While conversation is in progress, the concerned MSC automatically transfers the call to a new available channel under the new base station. The MSC is an important node which oversees the handoff operations. There are two phases of handoff—*initiation phase* and *execution phase*. In the initiation phase, handoff decision is taken, and the execution phase includes the actual handoff process of allocating radio resources through the exchange of control messages. The system designer must specify the minimum signal level at which an acceptable voice quality could be maintained. Then the handoff is triggered at a slightly stronger level called the *threshold signal*. The difference between the threshold signal and the minimum acceptable signal needs to be optimised to avoid unnecessary handoffs.

Depending on the information used and action taken for decisionmaking process, handoff can be divided into three strategies:

1. **MCHO:** Mobile Controlled Handoff, the mobile device initiates the handoff process, needs more complex terminal.
2. **NCHO:** Network Controlled Handoff, analysing the signals from the base stations; MSC can decide the best candidate base station for the handoff. The mobile node plays a passive role in that case.
3. **MAHO:** Mobile Assisted Handoff, it is a sort of NCHO for GSM networks. The mobile node itself assists the network to handle handoff by analyzing signals obtained from the neighbor base stations and sends it to MSC for handoff decision.

In all the above cases, the objective of handoff is to have low hand-off latency, low loss, efficient resource utilization and less number of control signaling. Knowledge of mobility information is very critical for the design of handoff algorithms.

Parameters Related to Handoff in Cellular Environments Consider a situation when a mobile node moves around the hexagonal cellular environments. As the mobile node moves, it generates handoff. The rate of handoff depends on many parameters. In a gross sense, it depends on the number of cell boundaries crossing per cell per unit time. So, the area of each cell and area of the cluster size determine the handoff rate. Apart from that, the mobility model and the traffic condition on the network play a great role on the handoff [14,15]. Several parameters that influence the hand-off rate are:

1. Population density of mobile nodes (number of mobiles /km^2) = ρ
2. Cell radius (hexagonal) in km = R

3. Number of cell within a cluster = N
4. Speed of the mobile (km/h) = V
5. Traffic load per mobile station in Erlang/mobile = λ
6. Number of cell boundaries crossing per unit time = μ_{cell}
7. Number of cluster boundaries crossing per unit time = $\mu_{cluster}$
8. Percentage of powered stations = α
9. Percentage of active mobile nodes among the powered stations = β

Let us consider the simplest mobility model where the mobile nodes are uniformly distributed on the cellular environment and move with a constant velocity V, and direction of travel relative to the boundary is uniformly distributed over the region (0 to 2π). Considering the fluid-flow model, the cell-boundary crossing rate μ_{cell} can be given as follows:

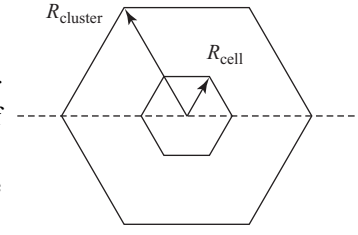
$$\mu_{cell} = \text{Population density of the mobiles} \times \text{Velocity of the mobile} \times \text{Perimeter of the cell} / \pi$$
$$= \rho V L_{cell} / \pi = \rho V 6R / \pi \quad (\text{where } L_{cell} = \text{perimeter of the hexagonal cell} = 6R)$$

Similarly, $\mu_{cluster} = \rho V L_{cluster} / \pi = \rho V 6 \sqrt{N} R / \pi$

As, $R_{cluster} = \text{Radius of the cluster} = \sqrt{N} R$

As, Area of the cell $A_{cell} = 2.598 R_{cell}^2$

Area of the cluster $A_{cluster} = N \times A_{cell}$



Now, consider that a cellular system is assumed to be a single cluster region. The total handoff generated per unit time due to the movement of the mobile across the cell boundary is

H_{total} = Number of cells × Number of cell crossing/time × % active mobile × % powered station × Traffic load on the mobile

Thus, $$H_{total} = N \times \mu_{cell} \times \beta \times \alpha \times \lambda \quad (13.1)$$

H_{total} includes the handoff rate due to cell boundary change as well as cluster boundary change. The total handoff rate due to change in cluster, which is known as inter-cluster hand-off, can be defined by replacing μ_{cell} with $\mu_{cluster}$ as

$$H_{inter\text{-}cluster} = \mu_{cluster} \times \beta \times \alpha \times \lambda \quad (13.2)$$

Now, handoff rate due to the cell change within a cluster per user basis is called intra-cluster handoff and is given as

$$H_{intra\text{-}cluster} = H_{total} - H_{inter\text{-}cluster} = (N \times \mu_{cell} - \mu_{cluster}) \times \beta \times \alpha \times \lambda \quad (13.3)$$

At this point, it is to be mentioned that the radius of a cell plays a great role in the generation of hand-off rate. Thus, while cell splitting helps in increasing the capacity of a cell, the handoff rate will also increase at the same time. This is the cost to be paid for the enhancement of capacity.

Example 13.1 In a GSM cellular system, the total geographical area is considered as a single cluster. Within a cluster there are 19 cells. The radius of each hexagonal cell is 2 km. The mobile node's population density is 50,000 mobiles /km^2. If the number of active users at a time is 12%, the percentage of the powered station is 60% and the call arrival rate per mobile is 0.05 Erlang, then find the following parameters for a mobile node moving with a velocity 5 km/h.

(a) Average cell crossing rate
(b) Average cluster-crossing rate
(c) Total hand-off rate
(d) Hand-off rate due to change of a cluster
(e) Hand-off rate due to change of cell within a cluster

Solution

$$\mu_{\text{cell}} = \text{Population density of the mobiles} \times \text{Velocity of the mobile} \times \text{Perimeter of the cell} / \pi$$
$$= \rho V L_{\text{cell}} / \pi = \rho V 6R / \pi \text{ (where } L_{\text{cell}} = \text{perimeter of the hexagonal cell} = 6R)$$
$$= 50000 \times 5 \times 6 \times 2 / 3.142 = 954805.8 \text{ crossing/cell/hour}$$
$$\mu_{\text{cluster}} = \rho V L_{\text{cluster}} / \pi = \rho V 6 \sqrt{N} R / \pi = \mu_{\text{cell}} \times \sqrt{N} = 4161902 \text{ crossings/cluster/h}$$

Total hand-off rate,

$$H_{\text{total}} = N \times \mu_{\text{cell}} \times \beta \times \alpha \times \lambda = 19 \times 954805.8 \times 0.12 \times 0.6 \times 0.05 = 65308.5/\text{h}$$
$$H_{\text{inter-cluster}} = \mu_{\text{cluster}} \times \beta \times \alpha \times \lambda = 4161902 \times 0.12 \times 0.6 \times 0.05 = 41982.8/\text{h}$$
$$H_{\text{intra-cluster}} = H_{\text{total}} - H_{\text{inter-cluster}} = 65308.5 - 41982.8 = 23325.7/\text{h}$$

13.4.4 Basic of Location Management

To save scarce radio resources, a group of cells (access points) can be defined as location areas (LA) and can only track in which location area the mobile node presently resides. The mobile node need not perform location update for every change of access point except the location area. During the packet delivery, the network tries to determine the exact location of a mobile and which cell it is in by the process of paging. A network may use multiple types of location areas. As an example, when the MN resides within a radio access network, the location areas is the number of radio cells, whereas in an IP network it could be IP subnets.

Several location update strategies are time based, movement based, distance based, parameter based and probabilistic update. Location update is made after a regular interval of time for time-based update. For movement-based update, the mobile performs a location update after traversing a predefined number of location areas. Most of the wireless networks like GSM, GPRS, 3GPP and 3GPP2 use a movement-based LA update with the threshold value of one. For distance-based update, a predefined distance threshold is used for LA update. The most realistic and dynamic LA update is the probabilistic approach where distance, movement or time thresholds are dynamically adjusted based on some probabilistic distribution of incoming call arrival.

Paging is an important part of the location-update process. When a network does not maintain precise location, it performs paging to determine its exact location. During the call delivery, the network warns the mobile by sending paging messages. When this message is sent to all cells under a location area, the method is called *blanket paging*. The network can search the mobile by sending multiple messages also with a paging delay cycle of more than one. But a key constraint is the time of paging that cannot be long enough to set up a call procedure. Upon receiving paging messages, the mobile needs to update its exact cell location. Another important thing is the paging area. Blanket paging should be avoided to save radio resources.

Location Update for 2G Networks For second generation GSM and IS-41 networks, location management (LM) consists of three parts—location update, call set-up and paging. Considering the basic 2G-network architecture in Fig. 13.12, the three steps can be explained.

The cluster of radio cells connected to MSC is called Location Area (LA). Within the LA, no location update is required while the mobile node is roaming within LA. But it requires LA update when roaming from one LA to other. The MSC is connected to the backbone network and the signaling network. It provides the coordination for location registration and call set-up.

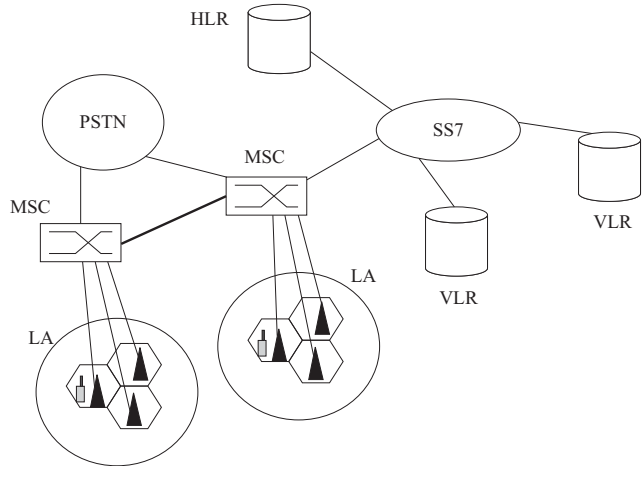


Fig. 13.12 *Two-tiered 2G network architecture*

The VLR is associated with each LA and maintains the location information of the visiting mobiles. The SS7 network carries out network signaling exchange operations using Common Channel Signaling (CCS). CCS provides simultaneous transmission of user data and signaling. CCS is implemented in a time division-multiplexing (TDM) format for serial data transmission.

While the mobile moves out of one LA to other, the beacon signal obtained from the Base Station (BS) is monitored, and the mobile node can sense that it has moved to a new location (new LA). The following steps happen for LA update:

1. The mobile node sends a registration request to the current BS that contains the mobile node's identification number (MIN).
2. The BS forwards the registration request to the MSC that generates the registration query to the VLR.
3. The VLR adds the entry of the registration record of the mobile and forwards the same to the HLR. The HLR information is obtained by translating the MIN.
4. The HLR performs authentication and security functionalities, updates the database for the new location obtained from VLR and sends an acknowledgement to the new VLR.
5. The HLR sends the cancelation message to the old VLR.
6. The old VLR removes the record of the mobile from its location database and returns a cancelation acknowledgement to the HLR.

The location update process is now complete and the HLR establishes association with the new VLR.

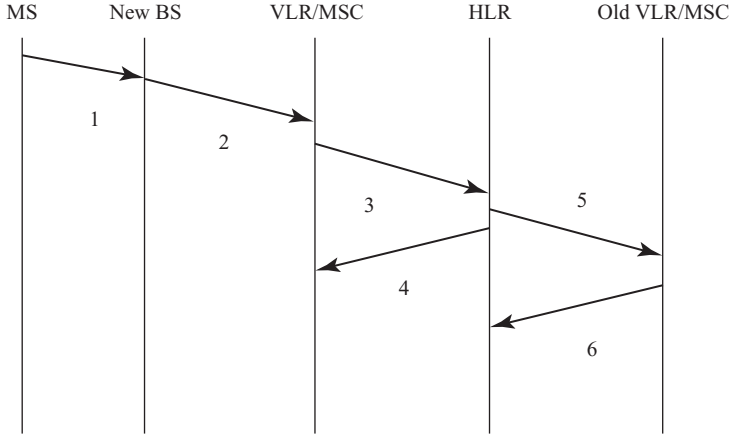


Fig. 13.13 *Location update signal flow graph*

Call Set-up Process When the mobile node is called, the network access point (BS) receives the request, and the network must locate the mobile node's position. The calling mobile sends the call request to the MSC via its serving base station. The MSC knows the location of the mobile through its MSC/VLR combination and in association with the HLR. The HLR sends the routes to the called MSC which in turn sends paging messages to get the

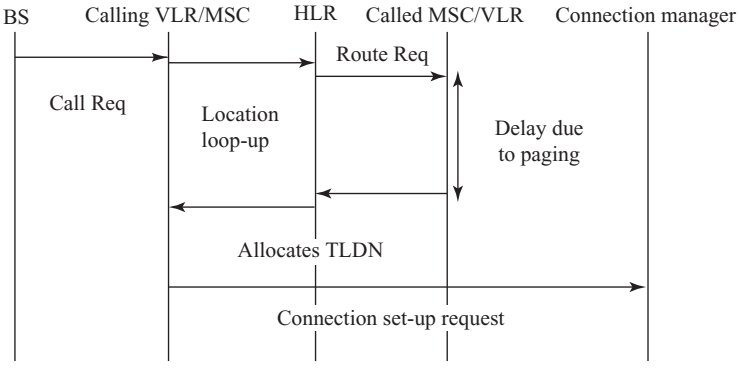


Fig. 13.14 *Call set-up procedure in 2G networks*

exact cell position of the called mobile. A temporary local directory number (TLDN) is allocated to the called mobile by the called MSC and is passed to the HLR. The HLR then passes it to the calling MSC. Using this TLDN, the calling MSC initiates a connection request to the called MSC through the SS7 signaling network. This call set-up procedure is illustrated in Fig. 13.14.

13.5 MOBILITY MANAGEMENT FOR 3GPP (UMTS) NETWORK

In Chapter 9, the mobility management procedure is already discussed for GPRS (2.5G) networks. In this section, we will discuss the mobility management in context of packet-switched (PS) domain for a 3GPP network. The key aspects of mobility management for 3GPP PS services are

1. Packet Mobility Management (PMM) contexts and states.
2. Location management and its interconnections with the management of the host-specific route between a mobile and its serving GGSN.
3. Changes of the Iu bearer service.
4. Handoff management.

PMM Context and States PMM context for the mobile is a set of information used by the network to track its location. The PMM context determines which network connections (bearers) between the mobile and the SGSN are to be maintained for the mobile and how the location of the mobile will be tracked. There are three major states of PMM context as described below.

1. **PMM-attach procedure:** This procedure allows the PS service domain of the network to be known. The PMM-attach procedure is required to be executed after the power-on of UE to get access to the network PS services. The location of UE is known in the SGSN and to the serving RNC. Unlike the GPRS system, the position of the mobile user is known up to cell level by the SGSN. Tracking is done by the SRNC up to the URA level. A packet-switched signaling connection is established between the UE and the SGSN during the PMM-connected state.
2. **PMM-detach procedure:** This allows the UE or the network to inform each other that the UE will not access the SGSN-based services. The UE is not known to the UMTS PS service and is not reachable to the network. In this situation, the UE may perform the attach procedure.
3. **PMM-idle:** In this state, the location of the UE is known in the SGSN to the accuracy of a RA. At this state, paging is required to know the position of the UE at the cell level. The mobile device performs routing area update if its RA changes.

While the mobile is in the PMM-attach state, the mobile's PDP context may have been created and activated. This is the case when the mobile node sends user packets over the PS-CN domain. During the state of transition of PMM-attach to PMM-idle, the mobile node's active PDP contexts will continue to remain in an active state on the GGSN and SGSN. It helps the mobile to come back from it's idle state to the attach state within a small time whenever the mobile nodes need to send/receive packets to/from over the PS CN domain. It also helps the PS-CN domain to support paging operations for the SGSN.

13.5.1 Location Management for PS Services

This can be defined with the help of Fig. 13.15. The RAN and the CN (Core Network) in a UMTS network use different location concepts to track the mobile nodes. RAN uses cell area or simply cell which is the geographical area served by one wireless BS and UTRAN Registration Area (URA)—URA is the area covered by a set of cells. Cells and URAs are used only in the RAN and are invisible in the CN nodes, whereas, LAs (Location Areas) and RAs (Routing Areas) are used for the CN.

LA is a group of cells used by the CS–CN domain to track the mobiles within the CS services. LA is handled by only one MSC/VLR. A globally unique identifier called Location Area Identifier (LAI) identifies each LA. Within one LA no location update is required with the CS–CN domain while the mobile is moving within LA.

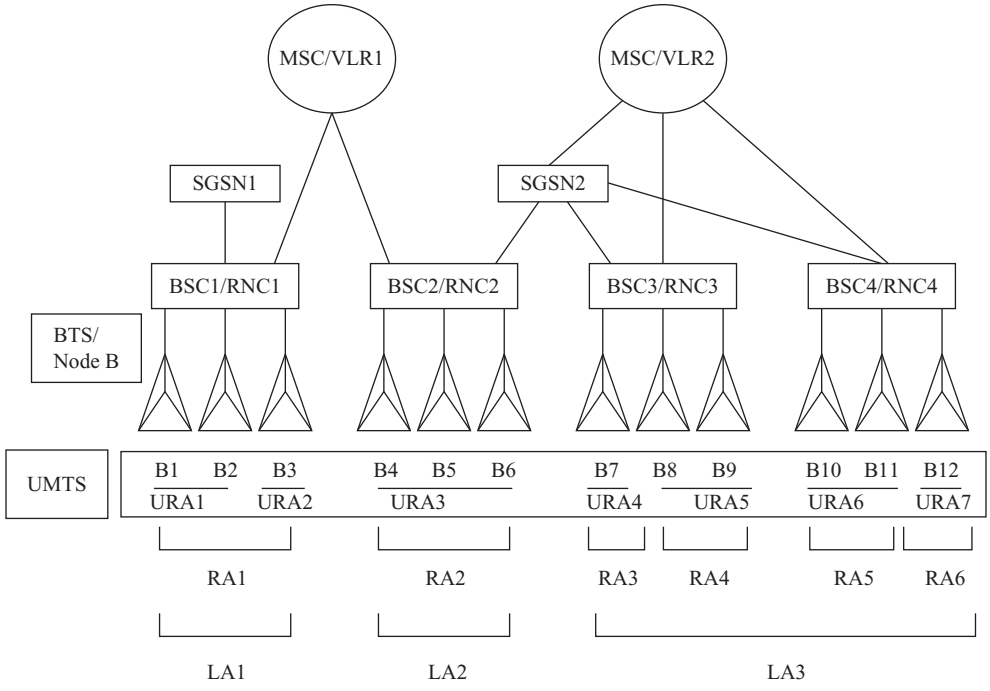


Fig. 13.15 *Concepts of LA, RA and URA*

RA is a group of cells used by the PS-CN domain to track the locations of mobiles that are using PS services. RA consists of one or more cells under RNCs that are connected to the same SGSN. One LA may contain more than one RA. A globally unique Routing Area Identifier (RAI) identifies each RA. The structures of LAI and RAI are shown in Fig. 13.16.

Structure of LAI

MCC (Mobile Country Code)	MNC (Mobile Network Code)	LAC (Location Area Code)

Structure of RAI

MCC (Mobile Country Code)	MNC (Mobile Network Code)	LAC (Location Area Code)	RAI (Routing Area Identifier)

Fig. 13.16 *Structure of LAI and RAI*

13.5.2 Location Tracking

For location tracking purposes, a mobile's activeness level is represented by the mode of RRC (Radio Resource Control) connection. The same RRC connection is used by the mobile to transport all signaling traffic and user traffic for its CS and PS services.

1. **RRC-connected mode:** When the mobile has established an RRC connection.
2. **RRC-idle mode:** When the mobile has not established any RRC connection.

During the **RRC idle state**, the mobile nodes can only be in the PMM-idle state because no signaling connection between the mobile and the SGSN can exist without the RRC connection. But when the mobile is in the RRC-connection mode, it may be either in PMM-connected mode or PMM-idle mode. So, location tracking is performed depending upon the mode of the mobile's RRC connection.

(a) When the mobile is in RRC-idle mode (PMM-idle) , location tracking is done by the SGSN at the RA level. The mobile will receive the Mobility Management (MM) system information broadcast by the RNCs at the RRC layer. MM information contains the cell and RA information in which mobile node currently residing. Upon receiving MM system information, the mobile node sends RA update toward the CN.

(b) When the mobile is in RRC-conneted mode, its location inside the RAN is tracked at the cell level by the RNCs. The mobile node is identified by a temporary identifier, called radio network temporary identifier, by the RNCs that is assigned dynamically.

During the RRC-connected state, the mobile node receives MM system information from the serving RNC over the established RRC connection and gets the information that it (mobile node) has moved to the new cell, RA or LA. If the mobile is RRC-connected and in the PMM-idle state then the SGSN knows the RA level. The mobile initiates the RA update procedure over PS-CN domain upon receiving the MM system information. If the mobile is in RRC-connected and PMM-connected mode then the SGSN knows the mobile node's serving RNC, and thus it can be tracked up to the cell posotion under the serving RNC. If the mobile changes its location to the new RNC position then the serving SGSN participates in the RNC relocation procedure to know the new position of the mobile under the new RNC.

13.5.3 Routing Area Update for UMTS Network

To ensure that the GGSN always knows where to forward user packets destined for the mobile, it requires an RA update. The mobile node invokes RA update when the mobile enters a new routing area, the periodic RA update timer expires, and reestablishment of RRC connection is needed.

RA update may be an intra-SGSN or an inter-SGSN update. With respect to Fig. 13.17, it is seen that intra-SGSN RA update occurs when the new RA and the old RA connect to the same SGSN, whereas, inter-SGSN RA update occurs when the new RA and the old RA connect to different SGSNs. Intra and inter-SGSN routing area update is discussed next.

Intra-SGSN RA Update Figure 13.17 describes the intra-SGSN RA update procedure. The mobile needs to be in the PMM-attach mode during the RA update process. If not then a change of state from PMM-idle to PMM-attach transition is required first.

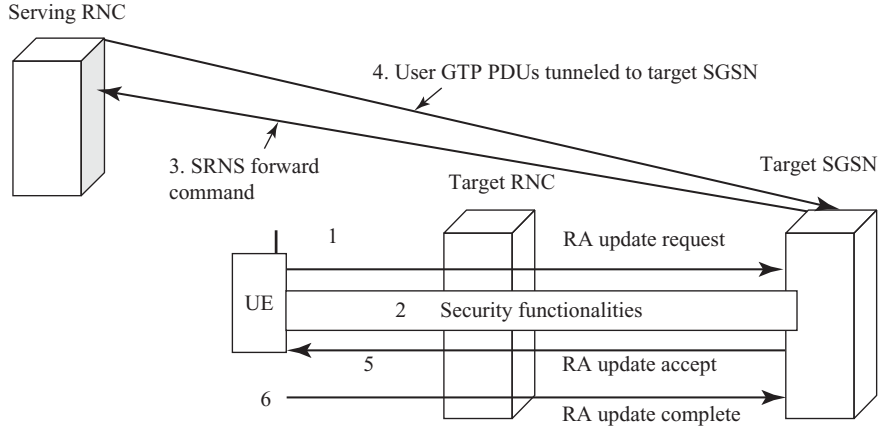


Fig. 13.17 *Intra-SGSN RA update for UMTS network*

The RA update request contains the information of P-TMSI, old RAI and old P-TMSI signature, update type and network capability. P-TMSI is assigned by the serving SGSN of the mobile node. From old RAI, the target SGSN will determine whether the intra or inter SGSN RA update is required. The P-TMSI signature is used by the SGSN to authenticate a P-TMSI. From the update type, the SGSN knows whether the RA update is triggered by a change of RA, a periodic RA update or a combined RA/LA update.

The following steps are performed for intra-SGSN RA update:

1. An RA update request is sent to the target SGSN via the target RNC. It triggers the establishment of Iu signaling in between the target RNC and the target SGSN.
2. The target SGSN authenticates the mobile as to whether the RA update request can be accepted by validating the mobile's P-TMSI. The SGSN has knowledge of the mobile's IMSI and the correct P-TMSI signature to validate P-TMSI.
3. Upon authentication, the SGSN updates the mobile's RAI, which it maintains for the mobile. The target SGSN sends the SRNS (Serving RNS) data forward command to the source RNS to tunnel the user traffic already buffered at this source RNC to the target SGSN.
4. The source RNC tunnels the traffic through the GTP tunnel to the target SGSN. The target SGSN then delivers the traffic to the destined mobile through the target RNC.
5. The SGSN sends the RS accept message to the mobile. The SGSN may also assign new P-TMSI.
6. The mobile responds to the RA update acceptance of the new P-TMSI by returning routing area update complete message to the SGSN that indicates the RA update complete procedure.

Inter-SGSN RA Update for UMTS Networks Inter-SGSN RA update is slightly different from intra-SGSN RA update because of the involvement of the target SGSN. Here too, the mobile node initiates the update procedure by sending the RA UPDATE REQUEST to the SGSN in exactly a similar fashion to intra-SGSN RA update, having the same information in the request frame. But one major difference is that the target SGSN needs to authenticate the mobile. To do this, the target SGSN first derives the source SGSN from the old RAI and P-TMSI carried in the RA update request message and then asks the source SGSN to validate the P-TMSI of the mobile by sending the SGSN context request message. It is to be mentioned that the SGSN context request carries the information of the old P-TMSI, old RAI and Old P-TMSI signature. Figure 13.18 illustrates the inter-SGSN RA update procedure.

If the mobile node is the authenticated one then the source SGSN will send the SGSN context response message that carries the mobile node's PMM context and PDP context. The PDP context provides the information to support packet delivery between the mobile and the network. The PMM context also provides mobility information about the mobile. The target SGSN initiates the process to update the PDP context on the mobile's GGSN during the RA update process. If the mobile is in the PMM-connected mode then the source SGSN could send the packets to the mobile before the RA update. The target SGSN should know the packet sequence number to be delivered after the RA update. This sequence number of the next packet to be sent, requested by the target SGSN, may be maintained by the source RNC. The source SGSN will send SRNS (Serving RNS) context request to the source RNC for this information. Receiving this request, the source RNS stops sending downlink packets to the mobile and returns the response of SRNS response message to the source SGSN by the sequence information.

If the authentication fails for P-TMSI then the source SGSN sends an error caused to the target SGSN. The target SGSN generates security procedure directly with the mobile node. If this authentication is negative, the target SGSN again rejects the request of RA update for this mobile. If positive, the target SGSN sends another SGSN context request message to the source SGSN to get the PMM and PDP contexts that carry the mobile's IMSI, old RAI and MS validation indicator. In reply, the source SGSN sends the PMM and PDP context of the mobile to the target SGSN or an error message if the souce SGSN does not have these information.

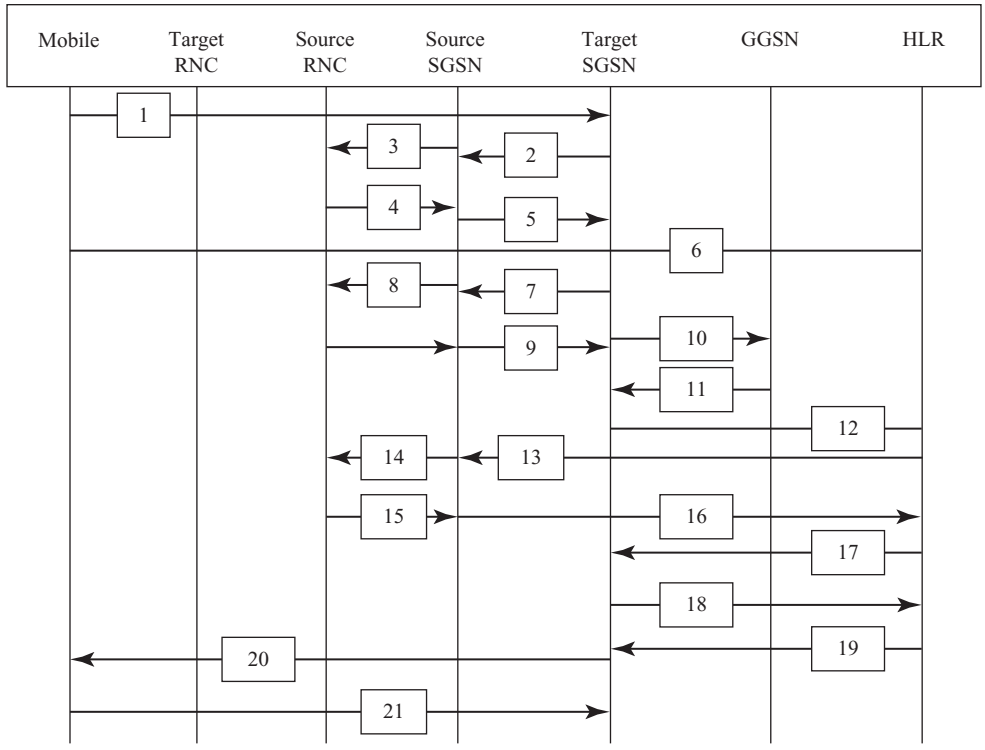


Fig. 13.18 *Inter-SGSN RA update process for UMTS networks*

The messages are described below.

1. Routing Area Update Request
2. SGSN Context Request
3. SRNS Context Request
4. SRNS Context Response
5. SGSN Context Response
6. Security Procedure
7. SGSN Context ACK
8. SRNS Data Forward Command
9. Forwards Packets
10. Update PDP Context Request
11. Update PDP Context Response
12. Update Location
13. Cancel Location
14. Iu Release Command
15. Iu Release Complete
16. Cancel Location ACK
17. Insert Subscriber Data
18. Inser Subscriber Data ACK
19. Update Location ACK
20. Routing Area Update Accept
21. Routing Area Update Complete

The target SGSN sends an SGSN context ACK message to the source SGSN. The target SGSN will initiate the process of PDP context update to ensure the serving GGSN in order to know in which SGSN the packets are to be delivered. The serving GGSN updates the PDP context for the new SGSN. The target SGSN will send this updated information to the HLR through the Gr interface that tracks the serving SGSN and thus the location update (LA) process is invoked. On receiving the LA update message, the HLR sends the location cancelation information to the source SGSN. The source SGSN removes the location and the serving-related information it has been maintaining. The source SGSN releases the Iu connection between the serving RNC and source SGSN. The HLR sends user services subscription information to the target SGSN, and in response it sends the Insert Subscriber Data ACK message to the HLR. HLRs send the Update Location ACK message to the target SGSN to indicate the completion of the LA update process.

Now, the target SGSN creates the PMM context for the mobile and sends the RA update accept message to the mobile. The target SGSN assigns a new P-TMSI for the mobile. The mobile sends the acceptance information to the target SGSN through the RA update complete message to the SGSN.

13.5.4 Serving Radio Network Controller (SRNC) Relocation for UMTS

Like GPRS, packets are routed between the mobile and the GGSN in UMTS. Figure 13.19 shows the routing path as an example.

A mobile can simultaneously transmit signals through multiple radio paths connected to different node Bs and these signals are merged to RNC1 known as the serving RNC. If the mobile moves to a different location during packet transmission, the routing path will change. As shown in Fig. 13.19, if the mobile moves towards Node B3 which is connected to RNC2, an Iur link between the RNCs is established so that the signal received by Node B3 can be sent to RNC1 via RNC2 (the Drift RNC -DRNC). RNC1 combines the signal from B2 and B3 and sends it to the SGSN1. The DRNC transparently routes the data through Iub and Iur interfaces and only performs Layer 1 and partial Layer 2 functionalities.

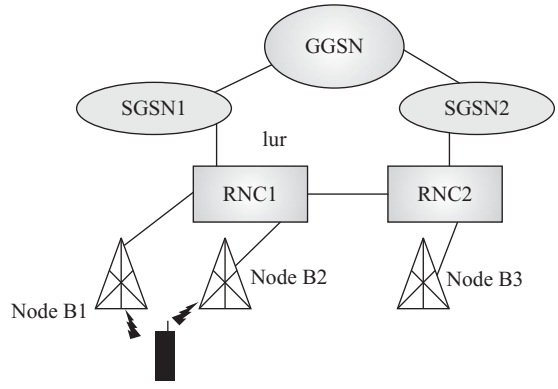


Fig. 13.19 *SRNC relocation for UMTS PS domain (initial steps)*

Now, suppose the mobile moves to the Node B3 position under RNC2. Now the routing path is Mobile → NodeB3 → RNC2 → RNC1 → SGSN1 → GGSN as shown in Fig. 13.20. This route is lengthy, and it does not make any sense to transmit through RNC1. So, RNC relocation is necessary and it removes RNC1 from the routing path.

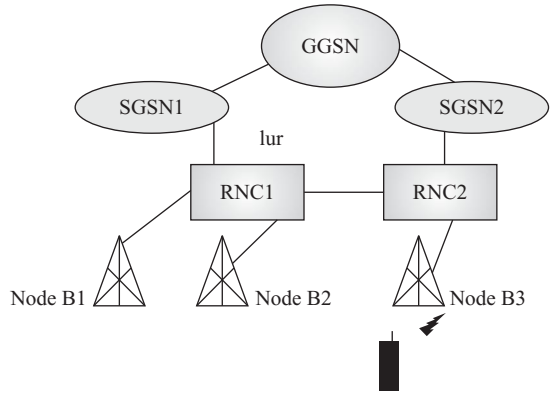


Fig. 13.20 *SRNC relocation for UMTS PS domain (MN moves Node B2-Node B3)*

After the RNC relocation, packets must route to GGSN via RNC2 → SGSN2 so that RNC2 becomes SRNC. At this point, hard handoff is invoked to connect the mobile from Node B2 to Node B3. SRNC relocation procedures for PS and CS are different. We will discuss the PS domain SRNC relocation procedure here as illustrated in Fig. 13.21. The mobile must be in PMM-connected mode.

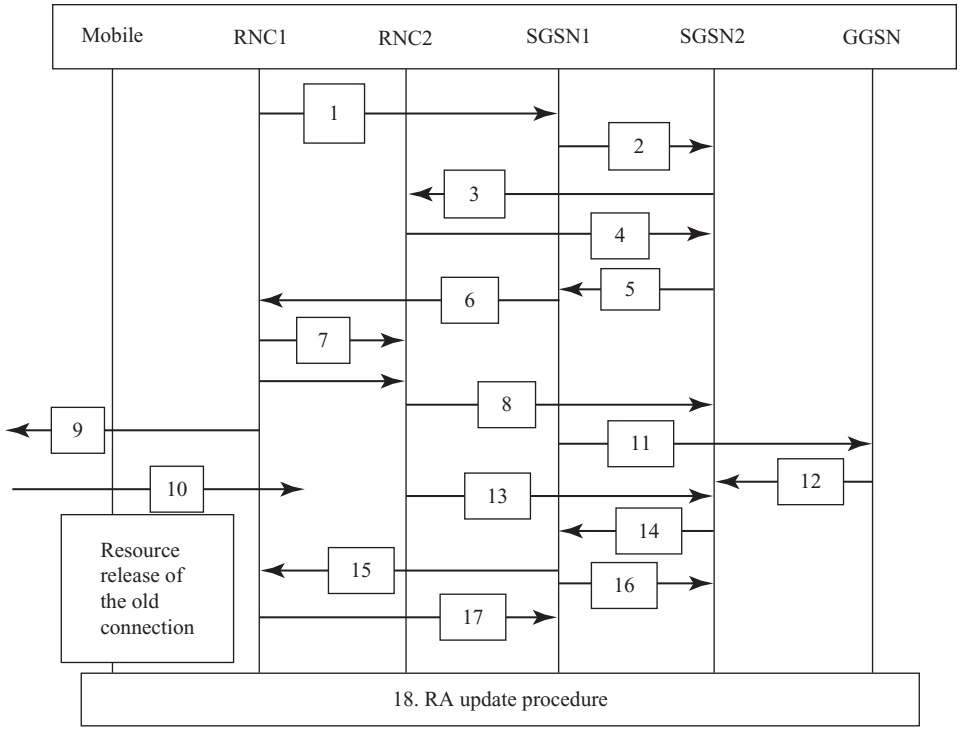


Fig. 13.21 *SRNC relocation procedure for UMTS PS domain*

Steps for SRNC Relocation

1. The RNC1 informs the SGSN1 through relocation required message. If both RNC1 and RNC2 are under the same SGSN, then Steps 2–5 are skipped.
2. For inter-SGSN RNC relocation, SGSN1 sends the MM and PDP contexts of the mobile to the SGSN2 through the forward relocation request message.

3–4. SGSN2 and RNC2 exchange the relocation request and response message pair to establish the Iu user plane transport bearers between the SGSN2 and RNC2 and also exchange routing information for packet delivery.

5. The forward relocation response is sent to SGSN1 by SGSN2 to indicate that SGSN2 and RNC2 are ready to accept packets buffered at RNC1.
6. SGSN1 sends the relocation command message to RNC1 for packet forwarding of buffered downstream packets to RNC2.
7. Upon receiving the relocation command message, the RNC1 starts the data forwarding timer and sends the relocation commit message to the RNC2 that contains the sequence information about the buffered packets to be tunneled to RNC2. At this point, RNC1 stops packet forwarding to the mobile obtained from GGSN.
8. RNC2 sends the relocation detect message to SGSN2 to let SGSN2 know that RNC2 is the SRNC. The core network switches the packet routing path from RNC1 to RNC2.

9–10. RNC2 and the mobile exchange information to identify the last up-stream packets received by the RNC2 and last downstream packets received by the mobile by exchanging the **RNTI reallocation and complete** message pair. RNTI reallocation contains information about LA, RA and RRC. The mobile node triggers RA update at the stage 18 because the RA has been changed. The mobile sends the RNTI reallocation complete message to the RNC2 after the reconfiguration.

1. Relocation Required
2. Forward Relocation Request
3. Relocation Request
4. Relocation Response
5. Forward Relocation Response
6. Relocation Command
7. Relocation Commit
8. Relocation Detect
9. RNTI Reallocation
10. RNTI Complete
11. Update PDP Context Request
12. Update PDP Context Response
13. Relocation Complete
14. Forward Relocation Complete
15. Forward Relocation ACK
16. Iu release Command
18. Iu release Complete

11–12. The SGSN2 switches the routing path from RNC1 to RNC2. The SGSN2 and the corresponding GGSN exchange the update PDP context request and response messages to modify the GGSN address, SGSN Tunnel End Identifier (TEID) and QoS profiles stored in the GGSN PDP context. This operation switches each GGSN connection from SGSN1 to SGSN2.

13. The RNC2 sends relocation complete message to SGSN2 and invokes resource release of the old Iu connection.

14–15. The SGSN2 instructs SGSN1 to release the old lu interface by exchanging the forward relocation complete and ACK messages.

16. The SGSN1 sends Iu release command message to RNC1.
17. The RNC1 returns the Iu release complete message to SGSN1 in reply to the data forward timer set up in Step 10.
18. Then the RA update procedure is initiated for inter-SGSN as described in the previous section.

13.5.5 Inter-RNC Hard Handoff

This is to be noted that for combined hard handoff with the SRNC relocation procedures, a similar message flow as descrided in Fig. 13.21 follows with some modifications. Inter-RNC hard handoff occurs when there is no implementation of Iur interface. Only the source RNC can initiate the inter-RNC hard hand-off process by measuring the radio channel quality and its knowledge of RAN topology. The combined hard handoff with SRNC relocation occurs at the beginning, when the RNC1 decides that the mobile should be involved and the mobile reconfigures the physical channel immediately after the SGSN1 sends the relocation command message to RNC1. In this process, the RNTI reallocation is not needed. The SRNC context of RNC1 is forwarded through the path of SGSN1 to SGSN2 to RNC2.

13.6 LIMITATIONS OF CURRENT TCP/IP NETWORKS FOR MOBILITY SUPPORT

The traditional TCP/IP was designed for fixed computer networks. The wireless access techniques only provide the mobility of homogeneous networks at the link layer, which is not appropriate for Internet mobility across heterogeneous networks. In general, the nature of network heterogeneity requires mobility support functions provided in higher layers. Besides, in mobile environments, the data link layer is based on wireless

access technologies (such as 3G, WLAN, etc.), which are characterized by low bandwidth, high bit error rates, faded and interfered signal with radio channel, etc. These wireless link features are encountered by the moving terminals, which may degrade the transport performance of high layers.

The basic problem with mobility in the Internet arises because an IP address has the dual role of routing identifier and endpoint identifier. Movement of a mobile node to a new position under a different subnet of the network invalidates the original IP address as a routing identifier, but the original IP address is still available through the DNS that is still valid as an endpoint identifier. In the mobile environment, the IP address of the MN has to be changed to represent the change of its point of attachment to the network when it moves from one network to another. In traditional TCP/IP, a change of the IP address makes it impossible for other devices to contact the device using a constant IP address. In addition, even if the device is able to obtain a new IP address dynamically, the transport connections established in the previous network will be broken after the change of IP address.

TCP congestion control is based on the assumption that the end-to-end path of a connection is relatively stable after connection establishment. In the mobile environment, the MN will change its access point of Internet attachment without notifying the TCP of its movement, and thus the existing end-to-end connection path has to be changed accordingly, which may violate this assumption and cause the TCP to make congestion control decisions based on invalid information [8].

Many applications based on traditional TCP/IP architecture are also limited in use in the mobile environment. For example, in Domain Name System (DNS), the Fully Qualified Domain Name (FQDN) is usually statically bound to a node's IP address. Thus the tight binding between the FQDN and the IP address will be invalid because of the dynamic change of IP addresses of the MN. When the MN moves to a new subnet, the subnet identifier portion of its IP address is no longer valid. Users who obtain a URL and use a DNS to look up the address will obtain an IP address that points to the original subnet, and not to the subnet where the mobile host is currently attached.

13.7 MOBILITY SOLUTION

The present era of computing has ushered in rapid growth of wireless networks with widespread use of portable computers and the popularity of the Internet. Seamless user mobility has become the need of the hour. Hence, networks of tomorrow have to be robust enough to deal efficiently with different mobility patterns of the users. Many protocols have been designed and implemented that support IP mobility. Among them, Mobile IPv4, MIPv6 and Hierarchical MIPv6 are noticeable. The solutions chosen by the IETF for managing mobility in the Internet, namely the Mobile IPv4 (MIPv4) [1], Mobile IPv6 (MIPv6) [3], and Location Independent Network Architecture for IPv6 (LIN6) [9], at the network layer protocol are very important.

MIPv4 is discussed in the previous chapter. It basically separates the routing identifier and endpoint identifier for the mobile host. The endpoint identifier is the original home network IP address assigned to the mobile and is called the *home address*. The home address is propagated into the DNS and is used by other hosts in the Internet to establish initial contact with the mobile host. The address is valid for software in conversation with the mobile host when the mobile node moves to a new subnet.

In MIPv4, when the MN is on its home network, it acts like any other fixed node (FN) of that network and requires no special mobile IP features. Each time it moves out of its home network and accesses a foreign network, it obtains a care of address (CoA), e.g., through Dynamic Host Configuration Protocol (DHCP), and informs its Home Agent (HA) of the new address by sending a registration request message to the HA. The HA is basically a router in the home network that forwards packets to the mobile node and must be aware of the care-of-address. The home-agent based architecture of Mobile IP is a lookalike of gateway-based 3G networks, where a gateway sits between the cellular network and the Internet. The gateway is responsible for routing and complex trafficking like charging, QoS and service provisioning, while the MIP HA protocol

does not include those under its original MIP protocol. The mobile node is responsible for maintaining the binding between the home address and care-of-address at the HA. When the mobile node moves to a different subnet, it sends a registration request to the HA containing binding between the HA and CoA. Upon receiving the registration request message by the HA, it replies to the MN with a registration reply message. The HA uses IP tunnels to forward packets to the mobile node.

13.8 ACCESSING EXTERNAL PDN THROUGH GPRS/UMTS PS DOMAIN

Figure 13.22 illustrates the situation for transporting data from the mobile node to any external packet data network via the GPRS/UMTS PS domain.

Before the mobile node accesses the external IP network, authentication and authorisation is a must. Two basic ways to access external IP network are *transparent access* where the GGSN does not play any role during the interaction between the mobile and IP networks, and *non-transparent access* [11] where the GGSN has active participation.

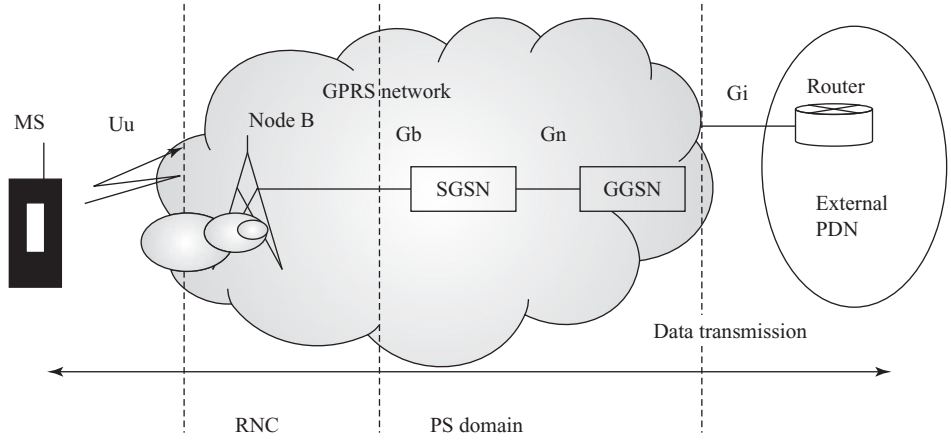


Fig. 13.22 *Data transfer to external PDN through UMTS PS domain*

13.8.1 Transparent Access

In the transparent access mechanism, the PS core network is used to send and receive IP packets over an external IP network. GPRS attach and activation of PDP context are the two important steps for transparent access through which the mobile node accesses the GGSN in the visited domain of the PS core network to send and receive data to/from an IP network. The mobile node acquires the IP address from the local PS domain as its PDP address, statically at the time of service subscription or dynamically during the activation of

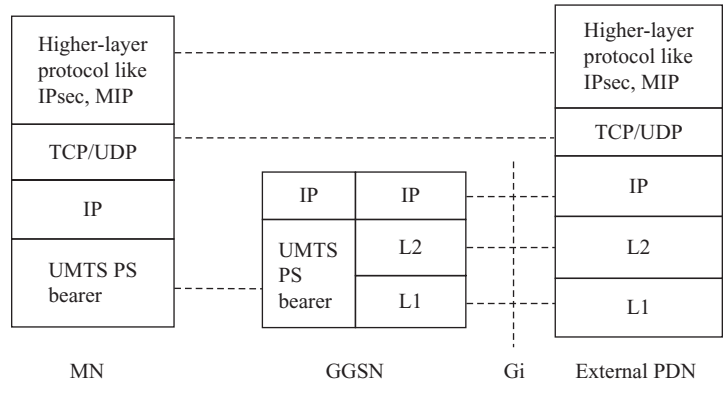


Fig. 13.23 *Protocol stack for transparent access to IP network via UMTS PS*

PDP context. The PS domain also checks its authenticity during the time of PDP context activation based on the user-subscription information. Obtaining an IP address from the local PS domain, the mobile node registers with the external IP network using IP based protocols. The local PS domain only serves as the carrier of the IP packets over the IP networks. IP is used as the packet data protocol over the PS domain between the GGSN and the external IP networks. Fig. 13.23 is the protocol stack for transparent access.

It is to be noted that for transparent access, each mobile needs an IP address from the PS-CN. It will require a large number of IP addresses as the number of mobiles is increasing day by day. The IPv4 address space may get exhausted and in future, IPv6-based solution may proliferate.

13.8.2 Use of Mobile IP for Non-transparent Access

As discussed in Chapter 10, Mobile IP operation needs two IP addresses. One is the mobile node's permanent home address and the other is the temporary care-of-address obtained from the visited domain of the mobile node. For non-transparent access, a mobile node uses Mobile IP home address as its PDP address for packet data forwarding through the UMTS PS core network. Each GGSN serves as an MIPv4 foreign agent and uses the IP address of the GGSN as its FA care-of-address (FACoA) at the local visited PS domain. In this way, the local domain need not provide any IP address. The normal MIPv4 registration process is implied when the mobile node visits the foreign domain. IP packets are addressed to the mobile node's home network and are routed using MIPv4 operation. It is the responsibility of the mobile node's home agent to tunnel the data to the mobile's current care-of-address of the visited domain which is basically the IP address of the GGSN. To operate in the non-transparent mode, it is required to impose MIPv4 functionality to the GGSN. The GGSN intercepts the IP packets, decapsulates and forwards the packets to the mobile along the path specified by the PDP context.

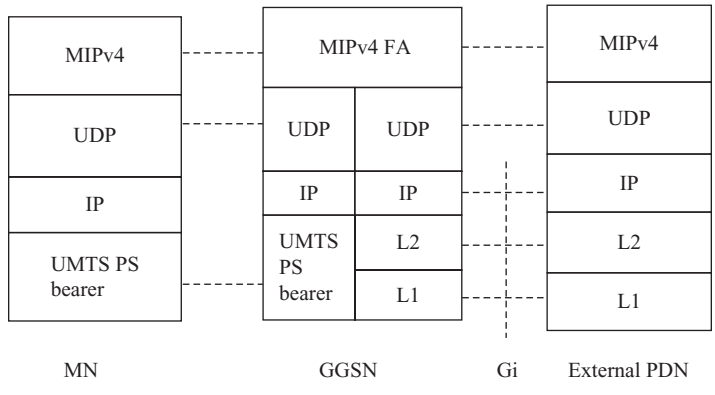


Fig. 13.24 *Protocol stack for non-transparent access using MIPv4*

The protocol stack for non- transparent access and the messaging required for MIPv4 operation is shown in Figs. 13.24 and 13.25 respectively.

GGSN learns the mobile node's home address from the MIP registration reply messages, and acts as the MIPv4 Foreign Agent. Within the local PS domain, every mobile node uses the FA CoA as the care-of address. Data packets addressed to the mobile node's home agent will be routed to the GGSN. To get access to the GGSN, the mobile node first initiates the PDP context activation with the address field of PDP context set to zero. By using the APN within the Active PDP context, the SGSN selects the correct GGSN that is again configured as the MIPv4 FA by the network operator. The create PDP context request is sent by the SGSN to the GGSN for the mobile keeping the PDP address set to zero.

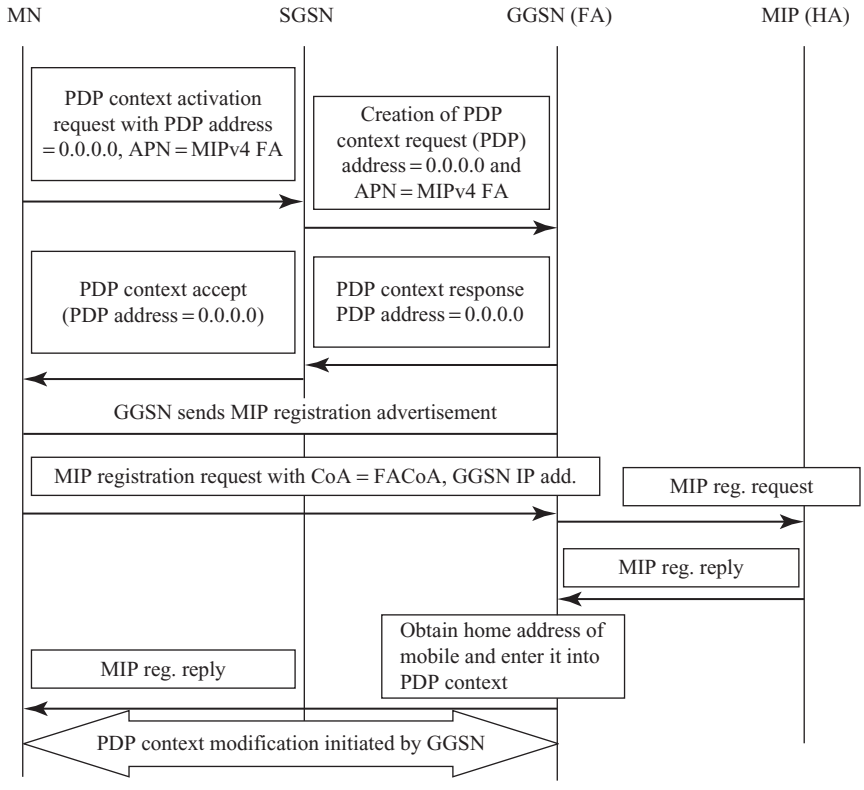


Fig. 13.25 *Non-transparent access of external IP network using MIPv4*

While accepting the activate PDP context, the GGSN (FA) sends the MIPv4 agent advertisement. Receiving the FA CoA, the mobile performs registration with the home agent HA by sending a registration request message. On receiving the registration request, the GGSN knows the home address or the Network Address Identifier (NAI) and maps them to the GTP-U tunnel identifier (TEID) for packet delivery to the mobile. FA on the GGSN will send the registration request to the HA of the mobile, and the HA in turn sends the MIPv4 registration reply back to the address of the GGSN (care-of address of the mobile). The GGSN extracts the home address information, puts it into the PDP context activation and initiates the PDP context-modification procedure for updating the PDP address on the SGSN.

13.8.3 Dynamical Accesses of IP Address from External Network

The Dynamic Host Configuration Protocol (DHCP) is used for getting an IP address from the external IP network with a DHCP server. Initially, PDP context for the mobile is activated without an IP address. After the active PDP context, the mobile node communicates with the external DHCP server to acquire an IP address. Before that, the PS-CN domain of the GPRS/UMTS networks relay DHCP messages between the mobile and the DHCP server. After assigning an IP address from the external network, the PDP context on the SGSN and GGSN is updated to include the mobile's IP address. Only the GGSN can initiate the updating process in the active PDP contexts on the SGSN and GGSN. The GGSN acts as a DHCP relay agent and transports DHCP messages over the PS-CN domain from the DHCP server before the mobile gets a valid PDP address. Now the GGSN learns about the IP address from the external DHCP servers based on the information of the APN obtained from SGSN during the PDP context activation. To do this, the DHCP relay agent on the GGSN intercepts and explores DHCP messages from the DHCP server to the mobile.

Figures 13.26 and 13.27 show the protocol stack and signaling flow messages for supporting IP address assignment using DHCP.

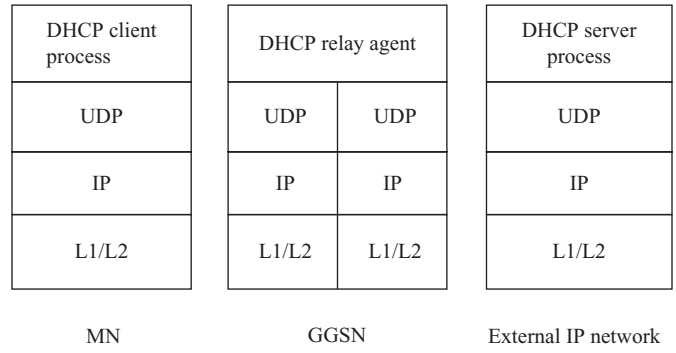


Fig. 13.26 *Protocol stack to access DHCP server through GGSN*

The mobile node sends the DHCPDIS-COVER message to the external IP network in order to acquire an IP address through the DHCP Relay Agent on GGSN, which gets information of the APN within the mobile's PDP context. On reply, the DHCP server sends DHCPOFFER with the IP address be assigned to the mobile. The mobile node may get multiple DHCPOFFER messages among which only the selected server mobile sends the DHCPREQUEST indicating that the mobile has accepted the IP address. The DHCP server sends the DHCPACK acknowledgement message that contains the entire configuration parameters assigned to the mobile node. The relay agent on GGSN extracts the assigned IP address of the mobile and passes it to the GGSN. The GGSN then puts this IP address to the active PDP context and initiates the PDP context modification procedure.

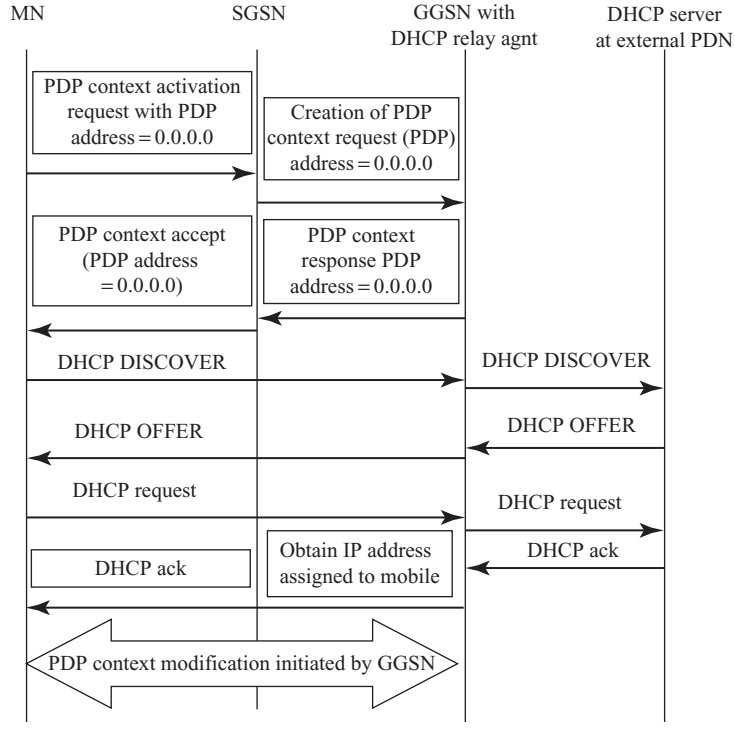


Fig. 13.27 *Signaling Flow messages for Acquiring IP address from external IP network using DHCP through GPRS PS domain*

Apart from all the described processes, the dialup connection may also be used to get access to external IP networks. PPP (point-to-point) is a most widely used protocol for setting up a link layer connection over a non-IP network. The most famous is the L2TP (Layer-2 Tunneling Protocol) standardised by IETF which extends a PPP connection over an IP network.

13.9 LIMITATIONS FOR MIP-BASED MOBILITY MANAGEMENT: USE OF MIPV6 AND ITS HIERARCHICAL MODELS

Because the traditional IP protocol has a variety of limitations for the next-generation Internet, the IETF defines a new network layer protocol, i.e., IPv6, attempting to replace the current IP protocol. The IPv6 is inherent in supporting Internet mobility management via MIPv6. MIPv6 follows the same basic principles as MIPv4, including home address, CoA, HA, and tunneling. The main difference is that the Foreign Agent (FA) no longer exists, and the security aspect has been improved. In addition, route optimisation has also been incorporated into MIPv6. Figure 13.28 shows the MIPv6 architecture and its operations. FA in MIPv4 gives two functionalities—provides care-of address to the visiting mobile nodes and helps the mobiles for movement detection.

In MIPv6, when the MN moves to another network, it acquires the CoA through either stateful or stateless address auto-configuration [Thomson and Narten 1998, 4]. By stateless auto-configuration, a mobile node can construct an address from its interface identifier and the subnet identifier advertised by the last hop router. It allows a mobile node to automatically configure a valid address on the local subnet without any additional support like FA in MIPv4. The last hop router is a standard IPv6 router with no modification. All the Access Routers (AR) are the IPv6 routers. When the mobile node discovers the last hop router and the subnet identifiers advertised by the router, it uses stateless auto-configuration.

After obtaining a new CoA, the MN registers to the HA and also to the correspondent node (CN) with Binding Update messages (BUs). The HA and the CN record this binding in the binding cache. After this, packets from the CN can be routed directly to the CoA of the MN with the CN's home address in the routing header. Similarly, the MN sends all packets to the CN directly using the home address destination option, which eliminates the triangular routing. In the event that the CN wants to communicate to the MN for the first time, the first packet is tunneled through the HA as in MIPv4. The HA intercepts any packets addressed to the MN's home address and tunnels them to the MN's CoA using IPv6 encapsulation. For discovering the HA, MIPv6 defines the Dynamic Home Agent Address Discovery (DHAAD) mechanism.

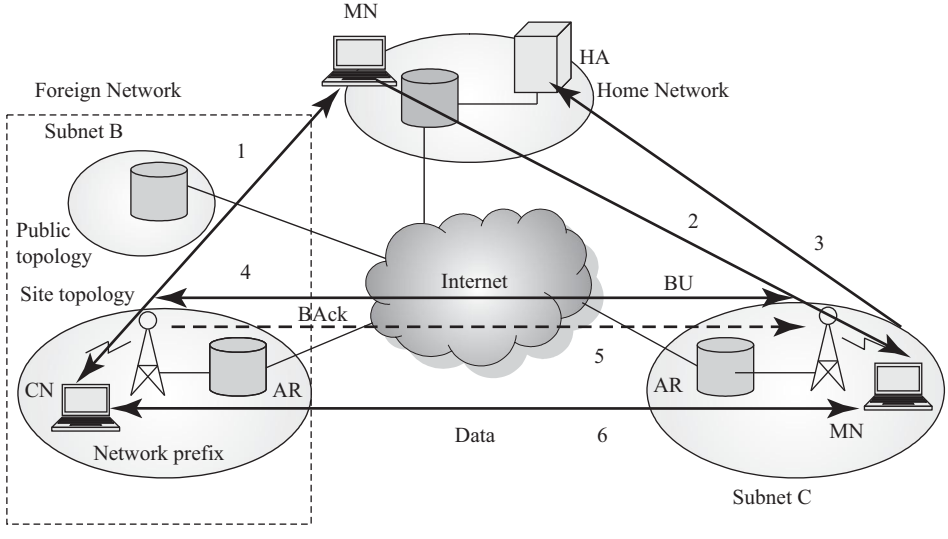


Fig. 13.28 *MIPV6 operations*

In MIPv6 (Johnson *et. al* 2004), route optimization is built into the base Mobile IP protocol. Route optimization allows the CN to send messages directly to the mobile node's care-of address and for the mobile node to reply using its care-of address as the source address bypassing the HA, thus removing the triangular routing problem in MIPv4. Route optimization is designed to be an integral part of the MIPv6. To support it, MIPV6 requires each IPv6 host and MIPv6 home agent to use a binding cache to maintain latest binding information received from the mobiles. When any IPv6 node wishes to communicate with another IPv6 node, it first checks for a binding update. If it exists, packets are sent directly to the destination's CoA, otherwise packets will be sent to the destination's home address.

In MIPv6, the MN is obliged to send a binding update message to its CNs and HA each time it changes its point of attachment. This causes significant processing overhead as the number of MNs increases. In addition, hand-off-speed performance is aggravated because the MN waits for an end-to-end path establishment so that it can receive packets on the new access router (AR). To overcome these limitations, IETF developed the Hierarchical Mobile IPv6 (HMIPv6) protocol to reduce overload and improve handover speed by separating the local mobility from global mobility. HMIPv6 proposes multi-level hierarchical network architecture. Hierarchical MIPv6 (HMIPv6) introduced a local entity within the access network, the Mobility Anchor Point (MAP), which can be located at any level in a hierarchy of routers, including the AR.

The idea is that the movement of the MN within MAP domain is not visible to the CNs and HA. So the later need not be notified of MN movements within the MAP's subnet. The MAP intercepts the packets destined for MN and tunnels them to the actual location of the MN. In HMIPv6, packets are destined to an address within the MAP's subnet and not to the MAP itself, and so the latter has to function as a typical HA for intercepting data packets as a proxy and tunneling them to the MN's location. An MN in a MAP domain is assigned two IP addresses—a Regional CoA (RCoA) on the MAP's subnet and an on-Link CoA (LCoA) that corresponds to the actual location of the MN. The later registers with the local MAP to provide binding information for its RCoA and LCoA. As long as the MN moves under the same MAP, it does not need to transmit binding updates to CNs and HA. Instead, it must locally inform the MAP of its movements, resulting in a change of its LCoA. When the MN moves to a new MAP domain, the former needs to register with the later as well as with its CNs and the HA. So the CNs and HA maintain the home address-RCoA binding whereas the MAP maintains the RCoA-LCoA binding and is unaware of the MN's home address. Figure 13.29 illustrates the HMIPv6 operations.

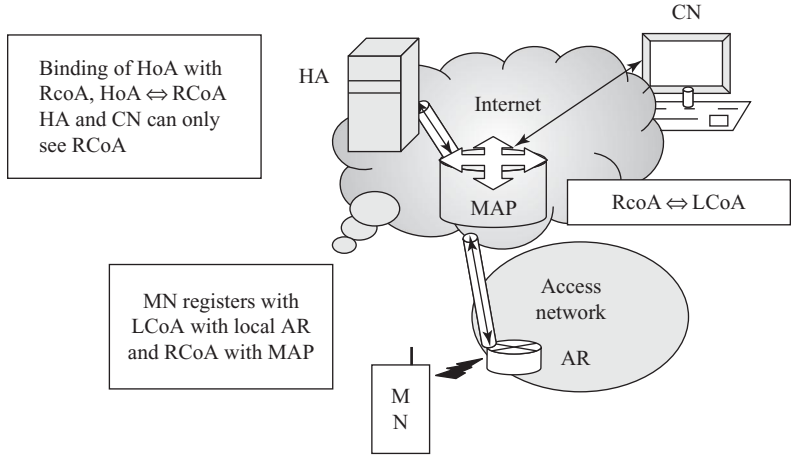


Fig. 13.29 *HMIPv6 architecture*

The global binding at the home agent is not changed until the mobile node moves outside the coverage area of its current MAP domain. The MN can perform route optimization at the CN, but only needs to do it

once, to establish a binding between the RCoA and HoA. CN can only see the HoA and RCoA but not the LCoA. The MN's location is not exposed except on the link between the MAP and the MN itself, where the LCoA appears in the tunnel header. This type of managing mobility is called localized mobility-management technique. It can improve efficiency in the binding update time and signaling overhead.

While many TCP applications are designed to cope with intermittent loss of connectivity by retransmitting unacknowledged packets, UDP applications will not be able to recover such losses. Furthermore, both TCP and UDP applications that rely on timely packet delivery within certain acceptable thresholds (e.g., VoIP and audio/video streaming applications) will be sensitive to the length of time an MN loses connectivity while performing handover. Such applications desire what is known as *seamless handover*; where seamless refers to handovers that are both the following:

1. **Smooth:** No (or very little) packet loss
2. **Fast:** Low latency

Fast handovers for Mobile IPv6 (FMIPv6) [7,8] is another proposal aiming at optimization for MIPv6. FMIPv6 attempts to acquire information that is needed to handover to a new link before disconnecting communication at the old link. It utilizes cooperating access routers which can request information from other access routers which are possible candidates for a handover. This is done by establishing a tunnel between the two access routers that allows the MN to send packets as if it was connected to its old access point while it is completing its handover signaling at its new access router. Therefore, it reduces the procedure time of movement detection, new CoA configuration, and binding updates, etc., during handover, and eliminates packet loss. Jung *et al*. [8] proposes a combination of both approaches of HMIPv6 and FMIPv6, which is designed to combine the advantages of both and provide additional improvements to reduce signaling overload, packet loss, and handover latency.

The problem of this hierarchical management is security, which is prone to attacks of a third party. To overcome route optimization and reverse routability, a security protocol in MIPv6 architecture is needed. But that may further increase signaling overhead for sending a binding update, particularly when the MN moves to a far point than its home network. Another problem of HMIPv6 routing is the single point of failure of a link, creating sudden service drop down. MAP contains bindings for all mobile nodes across a wide geographical area.

The problem arises from the fact that MIPv6 was not developed based on architectural consideration on providing mobility, but it was built as a simple extension of IPv6 from the viewpoint of routing. Mobility may also be defined in terms of two capabilities: (i) communication regardless of the location of the nodes, and (ii) continuation of communication even if a corresponding node changes its location.

Location Independent Network Architecture (LINA) for IPv6 LIN6 proposes an important alternative Internet mobility solution for the IPv6 protocol with the basic idea of separating the identifier and locator in the IPv6 address. LIN6 introduces the LIN6 ID for each node as the node identifier so that each node can be identified by its LIN6 ID, no matter where the node is connected and no matter how many interfaces the node has. In addition, it defines two types of network addresses—the LIN6 generalized ID and the LIN6 address. The LIN6 generalized ID is formed by concatenating a constant value called the LIN6 prefix before the LIN6 ID as shown in Fig. 13.30.

The concept of the ID-embedded locator of LINA is applied in IPv6. The locator is called LIN6 address. In IPV6, addresses are assigned according to the Aggregateable Global Unicast Address (AGUA) format as depicted in Fig. 13.30. LIN6 can use the same prefix as in AGUA on a foreign network for the upper 64 bits of a LIN6 address, and use a specially formed LIN6 ID for the lower 64 bits. The upper 64 bits of a given IPv6 address at the current interface address of a target LIN6 node are concatenated with the LIN6 ID of the target LIN6 node. The extraction operation simply draws out the lower 64 bits from a given LIN6 address as shown in Fig. 13.31.

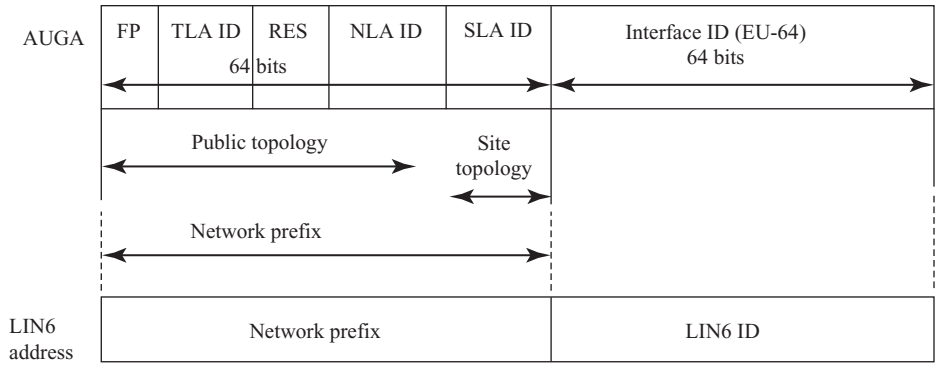


Fig. 13.30 *The packet format of AUGA and LIN6 address*

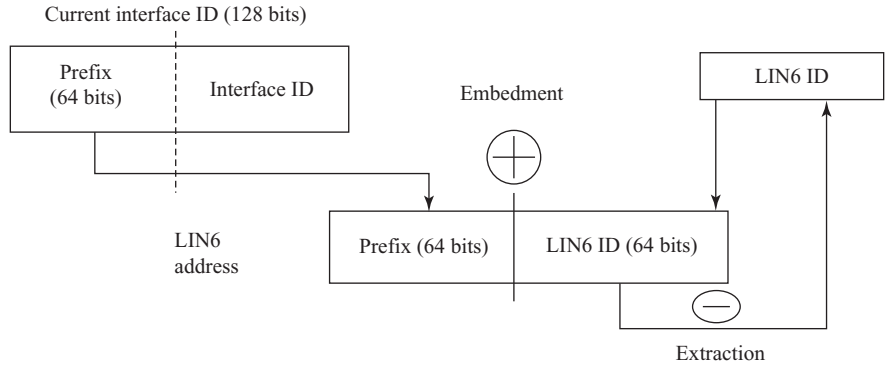


Fig. 13.31 *Embedments in LIN6*

LIN6 does not use DNS (Domain Name Server) for mapping the agent, but introduces a new dedicated server. The node on which the server runs is called mapping agent. To acquire the mapping of a target LIN6 node, a node first sends a query to DNS to obtain information on the designated mapping agent and then sends a query to any of these mapping agents to get the mapping. The result of mapping is stored for a lifetime within the mapping table. When a LIN6 node moves, it sends the mapping to one of the designated mapping agents of the node. A LIN6 node does not need to know its designated mapping agents in advance, as it can obtain the addresses of its designated mapping agents by querying the DNS. In MIPv6, the location of the HA is dependent on the address of the home network and needs to be placed within the home network even if there are multiple home agents. Placing of HA in a distributed way is difficult in an MIPv6 network. On the other hand, the location of a mapping agent in LIN6 is independent of the node identifier, and the designated mapping agent can be placed at any point on the network. Again, LIN6 mapping between the LIN6 generalised ID and the LIN6 address is performed only at the end nodes once the mapping is obtained from the mapping agent. It does not require any intermediate node processing for packet delivery and always guarantees end-to-end communication without using tunnels.

LIN6 performs two types of handoff, mapping update and mapping refresh request, to accommodate various securities functionalities. When an LIN6 node moves to a new location, it sends the new mapping to one of the designated mapping agents along with the correspondence nodes. This is called mapping update. The mobile node gets the correspondence nodes by inspecting the mapping table. The mapping refresh request is sent to the corresponding node to update its current mapping from the mapping agent and free it from any snooping attack. When a node gets a mapping refresh request message, it asks the mapping agent

and refreshes the new mapping. The handoff latency for LIN6 is less (~50 ms) compared to MIPv6 and is useful for voiceover IP application.

Summary

This chapter discusses the role of Internet Protocol (IP) for wireless networks. The details of packets routing mechanism and mobility management issues have been discussed in length for GPRS/UMTS networks. The aim of future generation networks is to be all-IP based along with heterogeneity. The traditional TCP/IP was designed for fixed computer networks. The wireless access techniques only provide the mobility of homogeneous networks at the link layer, which is not appropriate for Internet mobility across heterogeneous networks. So, IP must be replaced by Mobile-IP (MIP) based protocols for supporting global roaming across different networks. The limitations of MIP-based protocols have been discussed and as a solution strategy, the hierarchical version of MIPv6 and location independent network architecture have been overviewed for future generation network's mobility management.

References

[1] Perkins, C., *IP Mobility Support*, IETF RFC 2002, Oct. 1996.
[2] Deering, S. and R. Hinden, *Internet Protocol, Version 6 (IPv6) Specification, RFC 2460*, Dec 1991.
[3] Jhonson, D.B., C. Perkins and J. Arkko, *Mobility Support in IPv6, Internet-draft*, June 2002.
[4] Thomson, S. and T. Narten, *IPv6 Stateless Address Auto-configuration*, RFC 2462, Dec 1991.
[5] Deguang, L., F. Xiaoming and D. Hogrefe, *A Review of Mobility Support Paradigms for the Internet*, IEEE Communications Surveys and Tutorilas, 1st quarter 2006.
[6] Soliman, H. *et. al*., *Hierarchical Mobile IPv6 mobility management (HMIPv6)*, RFC 4140, Aug. 2005.
[7] Jung, H.Y. *et. al*., *Fast Handover for Hierarchical MIPv6 (FHMIPv6)*, Internet draft (work in progress), draft-jungmobileip-fastho-hmipv6-04, June 2004.
[8] Koodli, R., *Fast Handovers for Mobile IPv6*, RFC 4068, July 2005.
[9] Kunishi, M., *et. al*, *LIN6: A new approach to mobility support in IPv6*, Proc. 3rd Int Symposium, Wireless Personal Multimedia Communications, Nov 2000.
[10] Dixit, S. and Ramjee Prasad, *Wireless IP and Building the Mobile Internet*, Artech House, London, 2003.
[11] Cheng, Chen J. and T. Zhang, *IP-based Next Generation Wireless Networks*, J. Wiley, 2004.
[12] Bing, Lin Y. and A. Chun Pang, *Wireless and Mobile All-IP Networks*, J. Wiley, 2005.
[13] Saha, D., A. Mukherjee, I.S. Misra, M. Chakraborty, *Mobility Support in IP: A Survey of Related Protocols*, IEEE Network Magazine, Vol. 18, No. 6, pp. 33–40, November/December 2004.
[14] Park, S.J., Ji Young Song, J. Lee, K.J. Kim, and B.Gi Kim, *A Handover Scheme in Clustered Cellular Networks*, Future Generation Computer System, Elsevier, Vol 20, pp. 221–227, Feb 2004.
[15] Liu, Xin, and Weihua Zhuang, *A 3G/IP Interworking System Supporting Inter-cluster Soft Handoff, Wireless Personal Communications, Springer, pp.* 279–305, Sep 2003.

Questions for Self-Test

13.1 To carry data between the UE and the hosts on the Internet or other external PDN for UMTS/GPRS networks, mobile users use
a. IP as a transport mechanism b. GTP c. both mechanisms

13.2 Use of GPRS Tunneling Protocol ensures security from hackers.
a. True b. False

13.3 In UMTS network, a user transport plane can run on a different IP version than the user plane and they are independent.
a. True b. False

13.4 UTRAN and core network transport can theoretically run
a. on different IP versions
b. on the same IP version

13.5 The UMTS Radio Access Network for user traffic is based on
a. ATM Adaptation Layer 2 (AAL2)
b. ATM Adaptation Layer 5 (AAL5)

13.6 The UMTS Radio Access Network for signaling traffic is based on
a. ATM Adaptation Layer 2 (AAL2)
b. ATM Adaptation Layer 5 (AAL5)

13.7 Both UMTS Release 5 and 6 (R5 and R6) are evolved to accomodate all IP architecture.
a. True b. False

13.8 Packet Data Convergence Protocol is used to transport
a. point-to-point b. IPv4
c. IPv6 protocols d. all in above

13.9 The main function of PDCP is
a. header compression b. fragmentation

13.10 Gp interface is used between the SGSN and a GGSN
a. in a same PLMN b. in a different PLMN

13.11 CP is a transport layer protocol similar to UDP and TCP over an IP network, but SCCP runs over
a. ATM Adaptation Layer 5
b. Over AAL2

13.12 For implementing Mobile IP on GPRS networks, it is required that
a. GGSN always know the serving SGSN of the mobile node
b. GGSN may act as home agent or foreign agent

13.13 In UMTS there are two parts in the core network as
a. For CS connection _____ and GMSC
b. Fos PS connection _____ and GGSN

13.14 Mobile IP uses IP-in-IP _____.

13.15 SCCP protocol is similar to TCP but runs over _____ layer.

13.16 _____ is GPRS specific tunneling protocol.

13.17 _____ requires to get access to the new network when the user terminal moves from one to the new point.

13.18 Routing identifier and end point idenifier for the mobile node are separeted in _____.

13.19 GPRS _____ and _____ activation are the two steps of transparent access.

13.20 _____ access of getting IP address for PDP uses Mobile IP.

13.21 Mobile IPv6 eliminates _____ of MIPv4.

13.22 MIPv6 handles _____ aspect better than MIPv4

13.23 _____ does not use DNS (Domain Name Server) for mapping the agent, but introduces a new dedicated server.

13.24 MIPv6 requires each IPv6 host and MIPv6 home agent to use a _____ to maintain latest binding information received from the mobiles.

13.25 Placing of HA in a distributed way is difficult in _____network.

13.26 _____ attempts to acquire information that is needed to handover to a new link before disconnecting communication at the old link.

13.27 How do PDP addresses configure on mobile stations?

13.28 Describe the different methods of getting an IP address to access external PDN in a GPRS/UMTS network.

13.29 How does the coverage extend in a wireless network?

13.30 Consider two PLMN domains A and B under GPRS networks. Then explain the operation of data transfer when a mobile node moves from PLMN A to PLMN B.

13.31 With a pictorial representation explain the SRNC relocation for UMTS PS domain.

13.32 How is IP address from an external PDN dynamically accessed?

13.33 Describe the importance of GTP in relation to GPRS and UMTS networks.

13.34 Describe the packet routing and transport of user data in UMTS networks.

13.35 What is APN? How is it used in the packet data service of a GPRS network?

13.36 What are the different ways of getting an IP address for PDP context?

13.37 What are the challenges in seamless mobility in wireless envirionments?

13.38 What are the different types of mobility?

13.39 How does location management occur in UMTS PS-domain? Describe with a schematic diagram.

13.40 What are the different types of routing area update in UMTS networks? Describe the different steps of Intra-SGSN routing area update process.

13.41 How is an IP address aquired by the mobile node in a transparent access method?

13.42 How is Mobile IP used for non-transparent access PS domain ?

13.43 What are the basic limitations of Mobile IPv4?

13.44 How does hierarchical Mobile IP reduce signaling and handoff time within a domain?

13.45 Compare and constrast MIPv6 over LIN6 protocols.

13.46 In a cellular system the total geographical area is considered as a single cluster. Within a cluster there are 9 cells. The radius of each hexagonal cell is 1 km. The mobile node's population density is 30,000 mobiles /km^2. If the number of active users at a time is 14%, the percentage of powered station is 40% and the call arrival rate per mobile is 0.07 Erlang, find the following parameters for a mobile node moving with a velocity of 7 km/hr.

a. Average cell crossing rate

b. Average cluster crossing rate

c. Total handoff rate

d. Hand-off rate due to change of a cluster

e. Hand-off rate due to change of cell within a cluster

Ans. a. 401018.45 crossing/cell/h b. 1203055 crossings/cluster/h c. 14148/h d. 42443/h e. 28295/h

13.47 The hand-off rate is highly dependent on the cell radius and velocity of the mobile. Establish this statement by taking an example. Show how the cell radius affects the handoff if it becomes half of the original cell radius.

13.48 Draw a graph for different velocities starting from 1 to 6 km/h for total hand-off rate for the problem described in Problem 13.47.

13.49 In a cellular mobile system, the mobility information is always updated by HLR known as the location update. Within such system there are 100 cells each having a cell radius 2 km and a population density of 40,000 mobiles/km^2. Mobile nodes are uniformly distributed in a cell. The direction of movement of the mobiles are uniformly distributed over the region (0 to 2π). If the velocity of a mobile node is 5 km/h, calculate the total number of location update per hour.

Ans. 763844/h

13.50 Now consider the scenario of a UMTS cellular environment as depicted in Fig. 13.15 in this chapter. Calculate the number of LA, RA and URA updates due to the movement of a mobile.

Fundamentals of Wireless Local Area Networks

14

Introduction

Wireless local area networks (WLANs) are of the IEEE 802.11 standard for wireless local area networks. It works in similar way as the traditional LAN except for the wireless interface. Support for wireless local area networks (WLANs) in corporate offices and employee's homes is becoming a necessity for networking professionals, requiring new knowledge and training. WLAN has become very popular in college campuses, offices and in many public places as it supports wireless communication. WLANs provide high-speed (11/54 Mbps) data communication in small areas such as a building or an office and also connect users to IP networks, Internet or enterprise networks. The increasing number of PDAs and portable handheld devices have enhanced the popularity of WLAN, particularly in offices and hot-spot areas. It allows users to move around in a confined area while they are still connected to the network. WLAN uses the unlicensed ISM frequency band (Industrial, Scientific and Medical). The ISM band has a three-frequency range: 902-928, 2400-2483.5 and 5725-5850 MHz. The most popularly used WLAN standard is IEEE 802.11b which works at 2.4 GHz ISM band with a theoretical data rate of 11 Mbps.

There are several other IEEE802.11 WLAN varieties that define physical layers (PHY) and the Medium Access Control (MAC) layer for WLAN with a distinct purpose like security, QoS or increased data rate. Its early development included industry-specific solutions and proprietary protocols, but at the end of the 1990s these were replaced by standards, primarily the various versions of IEEE 802.11 (Wi-Fi–Wireless Fidelity).

The purpose of this chapter is to provide readers with a basic understanding of the 802.11 techniques, concepts, architecture and principles of operations.

14.1 IEEE 802.11

Architecturally, WLANs usually act as a final link between the end-user equipment and the wired structure of corporate computers, servers and routers. The standardisation of WLAN done by IEEE is known as IEEE 802.11 that describes the physical and data link layers. In 1997, the IEEE released 802.11 as the first internationally sanctioned standard for wireless LANs, defining 1 and 2 Mbps speeds [1, 2]. In September 1999, they ratified the 802.11b as high-rate amendment to the standard [3], which added two higher speeds (5.5 and 11 Mbps) to 802.11. The original 802.11 standard defines the basic architecture, features and services of 802.11b, with changes made only to the physical layer. These changes result in higher data rates and more robust connectivity.

14.2 WLAN TRANSMISSION TECHNOLOGY

The main transmission technology for WLAN is **spread spectrum** and **infrared**. **Frequency Hopping** (FHSS) and **Direct Sequence** (DSSS) modulation are the two methods used by spread spectrum transmission.

At the physical (PHY) layer, the IEEE 802.11 defines three physical techniques for wireless local area networks: diffused infrared (IR), frequency hopping spread spectrum (FH or FHSS) and direct sequence spread spectrum (DS or DSSS). While the infrared technique operates at the baseband, the other two radio-based techniques operate at the 2.4 GHz band. They can be used for operating wireless LAN devices without the need for end-user licenses. In order for wireless devices to be interoperable, they have to conform to the same PHY standard. All three techniques specify support for 1 Mbps and 2 Mbps data rates.

14.2.1 Frequency Hopping

The frequency hopping spread spectrum is analogous to FM radio transmission. The data signal is superimposed on, or carried by, a narrow band carrier that can change its frequency. The IEEE 802.11 standard provides 22 hop patterns, or frequency shifts, to choose from the 2.4 GHz ISM band. Each channel is of 1 MHz, and the signal must shift frequency or hop at a fixed hop rate. The signal hops from one frequency to another within a given frequency range. The transmitter device listens to a channel, and if it detects an idle time it transmits the data using the full channel bandwidth. If the channel is full, it hops to another channel and repeats the process.

14.2.2 Direct Sequence Modulation

This method uses a wide frequency band together with Code Division Multiple Access (CDMA). Signals from different units are transmitted at a given frequency range. A code is transmitted with each signal so that the receiver can identify the appropriate signal transmitted by the sender unit. The power levels of these signals are very low, just above background noise. Spread spectrum was first developed by the military as a secure wireless technology. It modulates a radio signal in a pseudo-random manner, and so it is difficult to decode.

14.2.3 Infrared Transmission

This method uses infrared light to carry information. There are three types of infrared transmission—diffused, directed and directed point-to-point.

1. **Diffused:** Infrared light transmitted by the sender unit fills the particular area such as an office. Therefore, the receiver unit located anywhere in that area can receive the signal.
2. **Directed:** Infrared light is focused before transmitting the signal. This method increases the transmission speed.
3. **Directed point-to-point:** Directed point-to-point infrared transmission provides the highest transmission speed. Here, the receiver is aligned with the sender unit. The infrared light is then transmitted directly to the receiver.

The light source used in infrared transmission depends on the environment. Light Emitting Diode (LED) is used in indoor areas, while lasers are used in outdoor areas.

14.3 SPREAD-SPECTRUM TECHNOLOGY

In spread spectrum (SS) technology [4], the signals from different sources combine into a single channel to fit into a larger bandwidth. It is useful to increase spectrum efficiency in wireless applications. In wireless transmission, the medium is air. All stations transmit and receive through the medium of air. So, the objective is to get such multiplexing schemes to sharing the radio channels and to make them interference free so that jamming could be avoided, and at the same time ensure that the transmission is free from the interception of any malicious intruder. This can be achieved by the spread-spectrum technology.

In this method, the original signal gets spread over a large bandwidth than the original bandwidth needed to transmit it. The signal is jumbled into the expanded bandwidth and thus becomes secured. Two basic steps are needed to implement SS.

1. The bandwidth allocated to each station is much larger than the required bandwidth.
2. If B is the original Bandwidth and B_{ss} is the BW for the expanded SS signal, the expanding of the original signal from B to B_{ss} must be done by a process that is independent of the original signal. The source signal and spreading process are independent. This is depicted in Fig. 14.1.

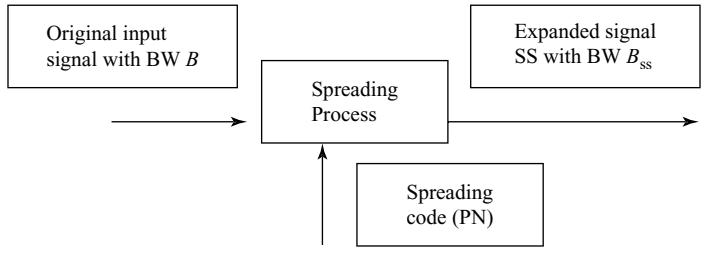


Fig. 14.1 *Spread-spectrum techniques*

The spreading code is the random pseudo-noise (PN) sequences that follow a pattern.

14.3.1 FHSS Technique

It uses M different carrier frequencies that are modulated by the source signal. At different time instants, the base signal is modulated with different hopping frequencies. After modulation, the bandwidth of the source becomes $B_{ss} > B$. If the number of hopping frequencies is M then the same BSS can multiplex M channels. The modulation technique used is called *Multiple Frequency Shift Keying* (MFSK). In FSSS, each station is allocated $1/M$ bandwidth but the allocation of the frequency is changed hop-by-hop. This is explained in Fig. 14.2.

The PN generator creates the N random bit patterns for each of the frequency-hopping periods T_{hop}. A frequency table is maintained in which M equally spaced frequencies are kept for each PN sequence bit pattern to be used for this hopping frequency and are passed on to the frequency synthesiser. It generates the appropriate carrier frequency of modulation of the source signal [4].

If there are many such randomly generated n-bit patterns and the frequency-hopping period is small then transmission security can be achieved both at the sender and receiver sites. The receiver needs to know the spreading signal sequence.

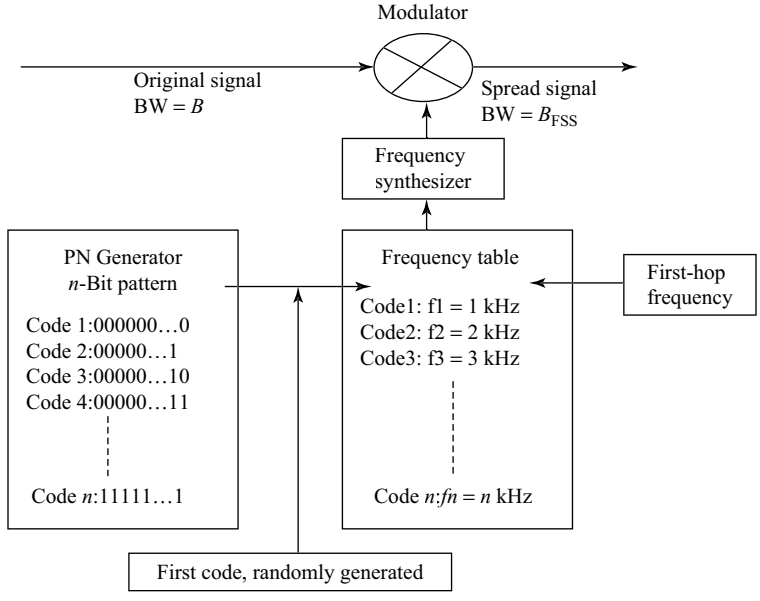


Fig. 14.2 *FSSS technique*

14.3.2 Direct Sequence Spread Spectrum

The DSSS can also be used to spread the bandwidth of the original signal. In this technique, each of the data bits is encoded with a spreading code having a chip rate of '*n*' bits. The DSSS works by taking a data stream of zeros and ones and modulates it with a second pattern, the chipping sequence. This sequence is also known as the Barker code, which is an 11-bit sequence (10110111000). The chipping, or spreading, code is used to generate a redundant bit pattern to be transmitted, and the resulting signal appears as wideband noise to the unintended receiver. One of the advantages of using spreading codes is that even if one or more of the bits in the chip are lost during transmission, the statistical techniques embedded in the radio can recover the original data without the need for retransmission.

A different line coding may be used for the original and chip generator, such as NRZ (not return to zero), Manchester and differential Manchester schemes [4]. To provide complete security, most spread-spectrum products include encryption. If the data rate of the original signal is ***n*** **bps** (bits per second) and the spreading sequence is 7 bits then the bandwidth required for the spread signal is 7 times, i.e., $7 \times n$ bps. If the spread signal from a different station cannot be combined and separated then the bandwidth cannot be shared. Figure 14.3 shows the DSSS technique.

The DS signaling technique divides the 2.4 GHz band into 14 numbers of 22 MHz channels, of which 11 adjacent channels overlap partially and the remaining 3 do not overlap. Data is sent across one of these 22 MHz channels without hopping to other channels, causing noise on the given channel. To reduce the number of re-transmissions and noise, chipping is used to convert each bit of user data into a series of redundant bit patterns called **chips**. The inherent redundancy of each chip, combined with spreading the signal across the 22 MHz channel, provides for error checking and correction functionality to recover the data. DSSS is primarily an inter-building technology, while FHSS is primarily an intra-building technology.

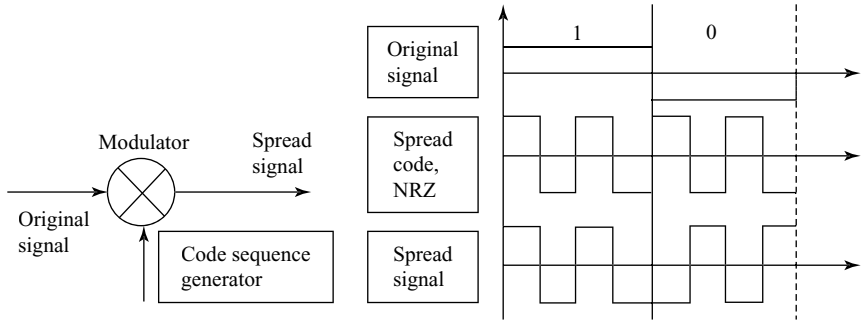


Fig. 14.3 *DSSS technique*

14.4 WLAN SYSTEM ARCHITECTURE

IEEE 802.11 supports three basic topologies for WLANs—the Independent Basic Service Set (IBSS), the Basic Service Set (BSS), and the Extended Service Set (ESS). All three configurations are supported by the MAC layer implementation. The system is constituted by the following entities:

1. **Station (STA):** The object of the communication, in general, a mobile station.
2. **Access point (AP):** A special central traffic relay station that normally operates on a fixed channel and is stationary—can be partially seen as the coordinator within a group of STAs. The AP is analogous to the base station in a cellular communication network.
3. **Portal (PO):** A particular access point that interconnects IEEE 802.11 WLANs and wired 802.2 LANs. Thus, it provides the logical integration between both types of architectures.

4. **Basic Service Set (BSS):** A set of STAs and eventually an AP, constitutes a Basic Service Set (BSS), which is the basic block of the IEEE 802.11 WLAN. The simplest BSS is constituted by two STAs that can communicate directly. This mode of operation is often referred to as an ad-hoc network because this type of IEEE 802.11 WLAN is typically created and maintained as needed without prior administrative arrangement for specific purposes. The BSS without an AP is a standalone network and cannot transmit data to other stations. This basic type of IEEE 802.11 WLAN is called Independent BSS (IBSS).

The second type of BSS is an **Infrastructure BSS**. Within an infrastructure BSS, an AP acts as the coordinator of the BSS. Instead of existing independently, two or more BSSs can be connected together through some kind of backbone network that is called the **Distributed System** (DS). The whole interconnected WLAN, some BSSs and a DS, are identified by the IEEE 802.11 as a single wireless network, called **Extended Service Set (ESS),** as shown in Fig. 14.4. An ESS is a set of two or more BSSs forming a single subnetwork. ESS configurations consist of multiple BSS cells that can be linked by either wired or wireless backbones called DS. The ESS appears as one large BSS to the Logical Link Control (LLC) to each STA. The DS is the backbone network that is responsible for the MAC layer transport of MAC service data units (MSDU). The DS could be 802.3 Ethernet, IEEE 802.4 token bus LAN, a fiber optic LAN or other wireless LAN networks. IEEE 802.11 supports ESS configurations in which multiple cells use the same channel, or use different channels to boost aggregate throughput.

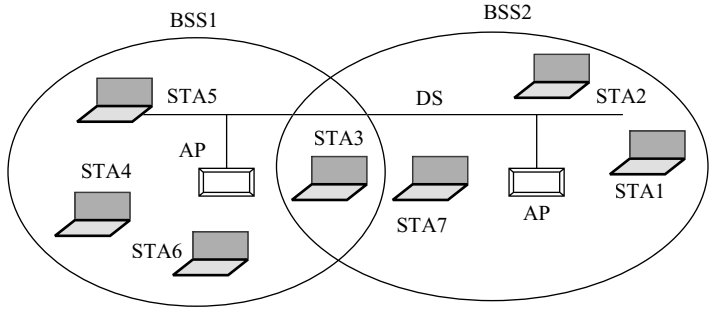


Fig. 14.4 *WLAN architecture: extended service set*

The association between an STA and a particular BSS is dynamic. The IEEE 802.11 does not restrict distribution system; it can be any IEEE LAN such as Ethernet. The STA within the ESS can be mobile or stationary. Stationary STAs are generally APs and a part of the wired LAN structure.

When the BSSs are connected together, two stations within the reach of each other can communicate without using AP. But when the stations are under different BSSs then they communicate through APs. The AP is necessary in order to perform a bridging function and connect multiple WLAN cells or channels, and also to connect WLAN cells to a wired enterprise LAN.

The IEEE 802.11 standard not only defines the specifications, but also includes a wide range of services:

1. Support of asynchronous and time-bound (time-critical) delivery services
2. Continuity of service within extended areas via a distributed system, such as Ethernet
3. Multicast (including broadcast) services
4. Network management services
5. Registration and authentication services

Such services are divided into those assigned to STA, called Station Service (SS), and to the DS, called Distribution System Service (DSS). Both categories of services are used by the IEEE 802.11 MAC sub-layer. The services assigned to the STA are:

1. Authentication/de-authentication
2. Privacy
3. MAC service data unit (MSDU) delivery to upper layer (IEEE 802.2 layer)

The services assigned to the DS are:

1. Association/disassociation
2. Distribution
3. Integration
4. Re-association

The BSSs are provided by all stations, including AP, conforming with IEEE 802.11, while the DS provides the DSSs.

The original 802.11 standard defines the basic architecture, features and services of 802.11b, with changes made only to the physical layer. These changes result in higher data rates and more robust connectivity.

14.5 IEEE 802.11 LOGICAL ARCHITECTURE

The logical architecture of the IEEE 802.11 standard defines the network's operation. It applies to each station consisting of a single MAC and one of multiple PHYs (IR, DSSS, OFDM, FSSS) as shown in Fig. 14.5.

Both the MAC and PHY layers include two management entities—MAC Layer Management Entity (MLME) and PHY Layer Management Entity (PLME). In this section, we will discuss the IEEE 802.11 MAC layer standard and its functionalities. The layout of the MAC layer in the presence of PHY is shown in Fig. 14.6. The Data Link Layer (DLL) is the combination of two sublayers—the Logical Link Control (LLC) sublayer and the Medium Access Control (MAC) sublayer.

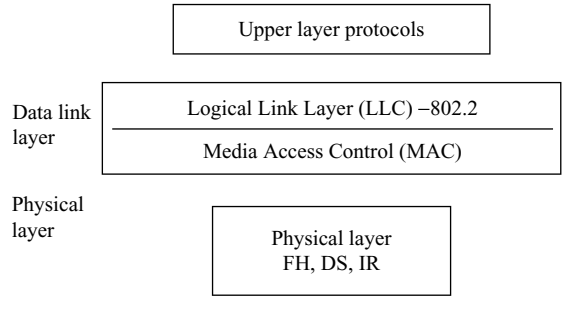


Fig. 14.5 *802.11 Standard on the bottom two layers of ISO model: PHY and MAC*

The goal of the MAC layer is to provide access control functions like addressing, access coordination, frame check sequence generation, and checking and LLC PDU (protocol data unit) delimiting for shared medium PHYs in support of the LLC layer. The 802.11 standard uses CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance), whereas standard Ethernet uses Carrier Sense Multiple Access with Collision Detection (CSMA/CD).

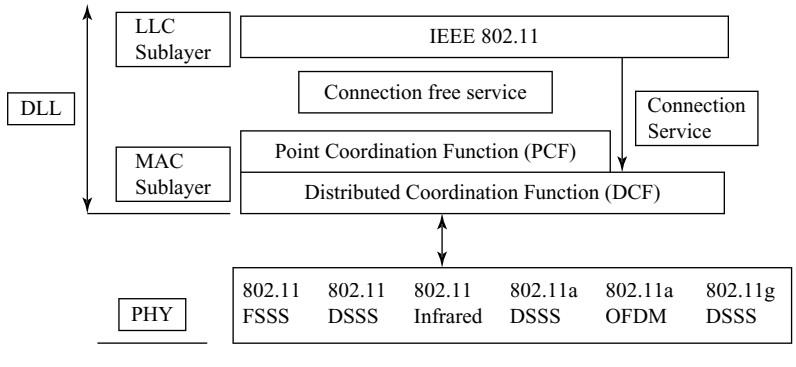


Fig. 14.7 *IEEE 802.11 MAC Layer Standard*

14.6 CARRIER SENSE MULTIPLE ACCESS WITH COLLISION DETECTION: CSMA/CD

CSMA/CD method is a very effective mechanism in a wired environment, enabling speeds of 10 (T-base), 100 (Fast-Ethernet), or 1000 (Gigabit-Ethernet). However, this mechanism allows collisions and supports exponential backoff mechanism, reducing the throughput in a very competitive environment with a high number of active users. Collision levels of 30–40%, even less, could cause a very significant degradation of the overall performance of the active users. There are several reasons why this method is not suitable for the WLAN standard.

1. Firstly, for CSMA/CD, a station must be able to send data and may receive collision signals at the same time. This can mean costly stations and increased bandwidth requirements.
2. Secondly, collision may not be detected because of the hidden station problem where the station hears the AP, but does not hear all other members of the cell.
3. Thirdly, the distance between stations can be great. Signal fading could prevent a station at one end from hearing a collision at the other end.

So, creating a mechanism to prevent collisions in the shared medium has always been a challenge for network designers. A different proposal was developed initially for wired media and later for wireless media based on collision avoidance techniques. The basic idea behind it is to listen to the media before data transmission and wait until the medium is free.

14.7 CARRIER SENSE MULTIPLE ACCESS WITH COLLISION AVOIDANCE: CSMA/CA

The 802.11 basic medium access allows interoperability between compatible PHYs through the use of the CSMA/CA protocol and a random backoff time following a busy medium condition. In addition, all the directed transmissions expect a positive acknowledgement frame (ACK) within a period of time. If not received, senders retransmit the frame. The 802.11 CSMA/CA protocol is designed to reduce the collision probability between multiple stations accessing the medium at the point in time where collisions would be most probable. The most probable collision may occur just after the release of any channel because multiple stations may wait for the free channels. So, the backoff algorithm is used to resolve medium condition conflicts. The basic access method for CSMA/CA is given in Fig. 14.7.

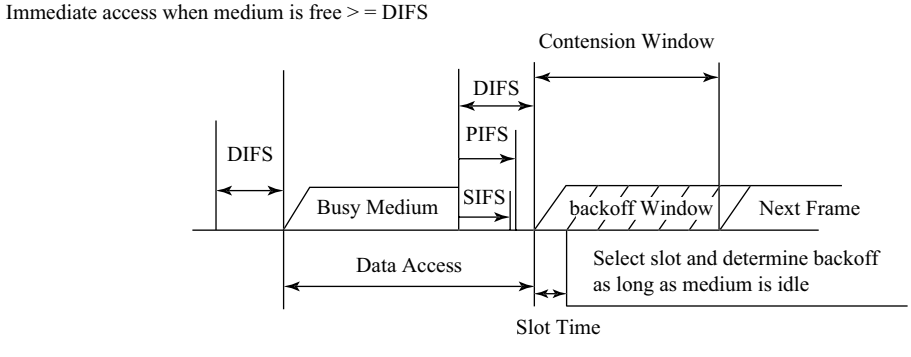


Fig. 14.7 *IEEE 802.11 Basic access method © Copyright IEEE 1997*

At the MAC layer, the 802.11 standard for CSMA/CA defines two different access methods—the Distributed Coordination Function (DCF) and the optional Point Coordination Function (PCF) which will be discussed

in the following sections after defining important related terms for inter-frame spacing.

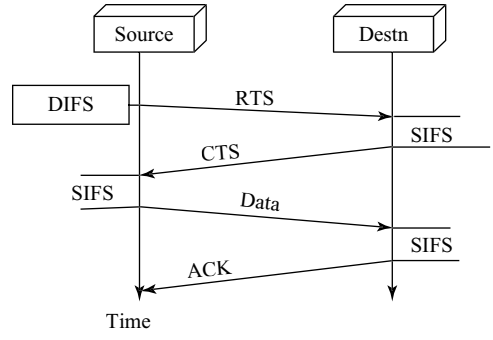


Fig. 14.8 *Carrier sense mechanism in IEEE 802. 11*

14.7.1 Inter-frame Spacing

There are three types of Inter-frame Spacing (IFS) for the original 802.11 MAC protocol as shown in Fig. 14.8.

SIFS (Short Inter Frame Spacing) is the smallest interval. This value for 802.11 PHY is fixed to 28 ms, time enough for the transmitting station to be able to switch back to the receive mode and be capable of decoding the incoming packet.

PIFS [(Point Coordination Function) (PCF) Inter-Frame Spacing] = SIFS + Slot time, is used by the access point (or point coordinator) to gain an access over the medium before any other station. The value of SIFS + Slot Time is 78 ms.

DIFS = Distributed Coordination Function (DCF) Inter-frame Space = PIFS + slot time. DIFS is the inter-frame space used for a station willing to start a new transmission, which is calculated as PIFS plus one slot time, i.e., 128 ms.

Priority access to the wireless medium is controlled through the use of the inter-frame space. The IFS intervals are mandatory periods of idle time on the transmission medium. Stations only required to wait an SIFS have priority access over those stations required to wait a PIFS or DIFS before transmission. SIFSs thus have highest priority access to the communication medium. The back-off timer is expressed in terms of the number of time slots. Figure 14.7 shows the basic access method for IEEE 802.11.

14.7.2 Distributed Coordination Function (DCF)

The distributed coordination function in 802.11 is based on a CSMA/CA mechanism. The DCF works by a station willing to transmit data sense in the medium first. If the medium is busy, the station defers its transmission to a later time, but if the medium is free for a specified time of DIFS, the station transmits. The receiving station then checks the CRC (Cyclic Redundancy Check) of the received packet and sends an acknowledgement (ACK) packet. Upon receiving this ACK, the transmitting station can understand that there were no collisions detected. If the sender does not receive an ACK then it re-transmits the last fragment. The DCF operates solely on an ad-hoc network mode or coexists with the PCF in an infrastructure mode.

Hidden stations do not hear each other but only hear an AP. As a result, there will occur collisions between stations during the transmit and receive process. To reduce the collision probability, the 802.11 standard defines virtual carrier sense mechanism.

14.7.3 Virtual Carrier Sense

A source station cannot hear its own transmission when a collision occurs and it continues transmitting the MAC Protocol Data UNIT (MPDU). For large MPDU, this is a wastage of bandwidth. The solution is RTS/CTS control frames. As shown in Fig. 14.8, a waiting station for transmission sends a frame called Request-To-Send (RTS) that contains a source and destination and the duration of the following transmission packets (the packets and ACK).

If the medium is free, the destination station responds with a response control packet called Clear-To-Send (CTS), which includes the same duration information. All stations receiving either RTS and/or CTS, set their **virtual carrier sense indicator called Network Allocation Vector (NAV)** for the given duration and use this information together with the physical carrier sense when sensing the medium. The NAV maintains a prediction of future traffic on the medium based on the duration information that is

announced in RTS/CTS frames prior to the actual exchange of data. So, stations receiving a valid frame will update their NAV with the information received in the duration ID field. In this way, all stations reserve the medium and inform the PHY layer that the medium is busy.

RTS and CTS control frames are relatively small compared to the maximum data frame size (2346 octets). RTS is of 20 octets and CTS is of 14 octets. The RTS control frame is first transmitted by the source station after successfully contending for the channel with a data or management frame queued for transmission to a specified destination station. All stations under a BSS, hearing the RTS packet, read the duration field and set their NAVs accordingly. The destination station responds to the RTS packet with a CTS packet after an SIFS idle period has elapsed. Stations hearing the CTS packet look at the duration field and again update their NAV. Upon successful reception of the CTS, the source station is virtually assured that the medium is stable and reserved for successful transmission [1].

Thus, for **medium reservation,** RTS/CTS frames are sent prior to the actual data frame. RTS and CTS frames contain information, on the duration ID field relative to the period that the medium is going to be reserved for transmission of actual data and its ACK frame. All STAs within the reception range of either the originating STA or the destination STA will learn of the medium reservation. Finally, the virtual carrier-sense mechanism makes use of network allocation vector.

The mechanism reduces the probability of a collision on the receiver area by a station that is hidden from the transmitter to the short duration of the RTS transmission, because the station hears the CTS and reserves the medium as busy until the end of the transmission. The duration information on the RTS also protects the transmitter area from collision during the ACK that is mainly caused from stations that are out of range of the acknowledgment station. Due to the short frames of RTS and CTS, the method also reduces the overhead of collisions. If the data is longer than RTS/CTS, it is possible to transmit data without RTS/CTS under the control of a parameter called RTS thresholds [5]. Figure 14.9 shows the mechanism of data transmission using RTS/CTS.

Without RTS/CTS, all stations hearing the data frame adjust the NAV based on the duration field value, which includes the SIFS interval and ACK following the data frame as shown in Fig. 14.10.

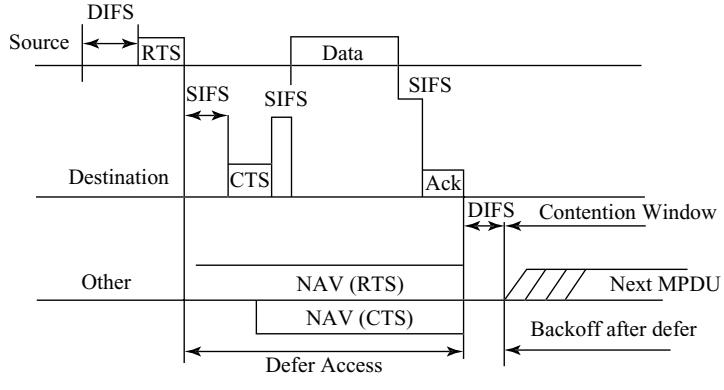


Fig. 14.9 *Data Transmission in 802.11 with RTS/CTS frames [1], © Copyright IEEE 1997*

Slot Time In DCF, the time immediately following an idle DFS is slotted (refer to Fig. 14.9) and the station is allowed to transmit only at the beginning of each slot time. The slot time is set equal to the time needed at any station to detect the transmission of a packet from any other station. It is again dependent on the PHY layer as given in Table 14.1 [6]. It accounts for the propagation delay, the receiving time from the receiving state to the transmitting state.

Table 14.1 *Contention Window size and slot time for different PHY layers*

PHY	CW_{min}	CW_{max}	Slot Time
FSSS	16	1024	50 μs
DSSS	32	1024	20 μs
IR	64	1024	8 μs

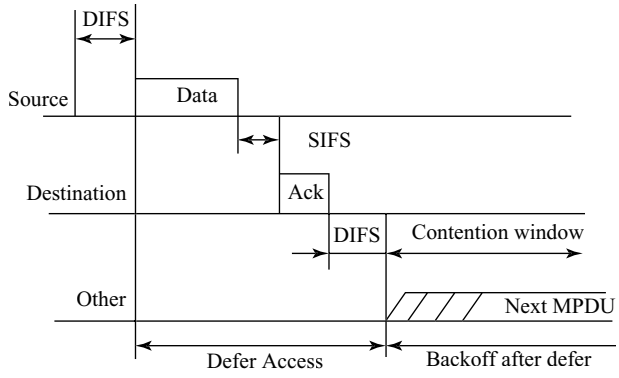


Fig. 14.10 *Transmission of MPDU without RTS/CTS [1], © Copyright IEEE 1997*

Back-off Algorithm DCF also adopts an exponential back-off scheme. Backoff is a well-known method used to resolve contention between different stations waiting to access the media. At each packet transmission, the back-off time is chosen uniformly between (0, CW-1), where CW is the value for the contention window. For the first transmission attempt, CW is set to CW_{min}, the minimum contention window. After each unsuccessful attempt, CW is doubled until it reaches the maximum value of CW_{max} [Qiang, 6] as shown in Fig. 14.11 within the range 7 to 63. The values of these two parameters depend on the physical layer and are defined in the standard. Table 14.1 shows the specific values of CW_{min} and CW_{max} with respect to PHY layers according to the IEEE 802.11 standard.

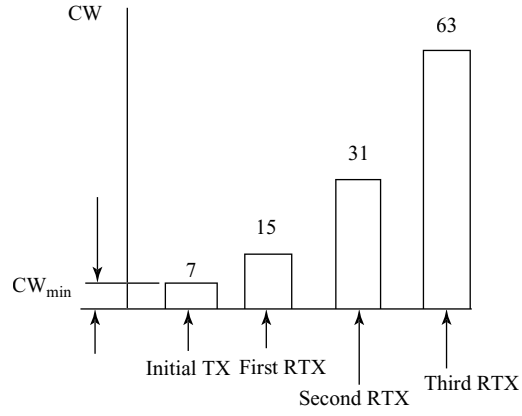


Fig. 14.11 *Exponential increase of backoff window*

Each station on BSS listens to the network. The first station begins the transmission with the allocated time slot needed to finish it. If another station hears the first station talk, it stops counting down its back-off timer. When the network is idle again, the station resumes the countdown. Each node starts a random backoff timer when waiting for the contention window. This timer counts down to zero while waiting in the contention window. Each node gets a new random timer when it wants to transmit. This timer is set on until the node has transmitted. The 802.11 standard defines an **exponential backoff algorithm** which must be executed when the station senses the medium before the first transmission of the packet; and the medium is busy after each retransmission and, after a successful transmission. When the medium has been free for more than DIFS and the station decides to transmit a packet, the backoff algorithm is not used.

14.8 MAC FRAME FORMAT AND FRAGMENTATION

The general MAC frame format is shown in Fig. 14.12. The IEEE 802.11 supports three different types of frame—management frame, control frame, and data frame.

1. **Management frames:** These are used for initial communication between stations and access points station association and disassociation with the AP, timing synchronisation, and authentication and deauthentication.

2. **Control frames:** Control frames are used for accessing the channel and ACK frames, for handshaking during the contention period (CP), for positive ACK during CP and to the end of the contention free period (CFP).
3. **Data frames:** They are used for carrying data and control information during CP and CFP.

Frame control (FC) is a 2-byte long field, which is again divided into many subfields as shown in Fig. 14.12.

(*a*) *Duration ID:* This field defines the duration in microseconds of the transmission that is used to set the value of the NAV. In one control frame, this field defines the ID of the frame.

The type field identifies the frame for management, control or data; whereas, the subtype field identifies the type of frame such as RTS or CTS.

(*b*) *Addresses:* There are four address fields each of 6-byte length. The meaning of each address field depends on the value of the To DS and From DS subfields.

(*c*) *Sequence control:* This field defines the sequence number of the frame to be used in flow control.

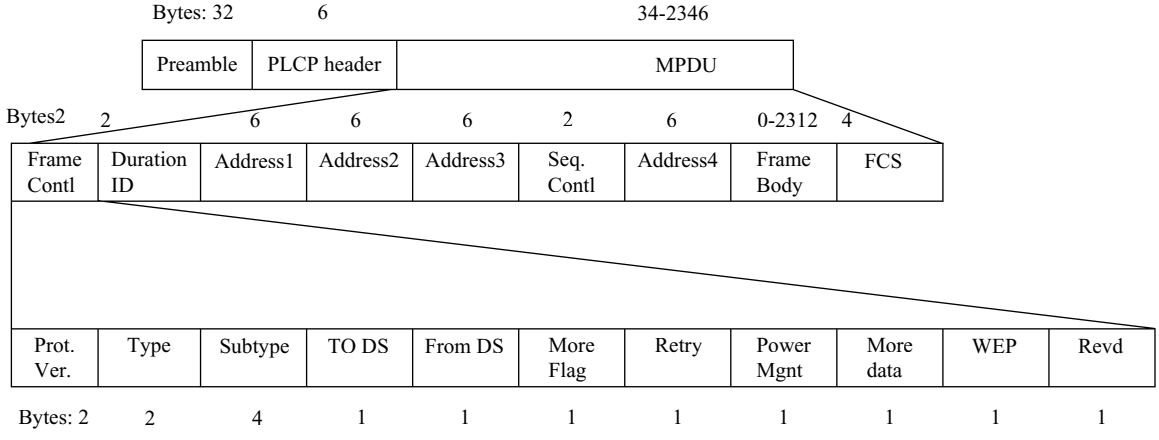


Fig. 14.12 *802.11 MAC frame format*

(*d*) *Frame body:* This contains information based on the type and the subtype defined in the FC field. It is the variable length field consisting of data payload and seven octets for encryption/decryption if the optional Wired Equivalent Privacy (WEP) security protocol is implemented.

(*e*) *FCS:* It is 4 bytes long and contains a CRC-32 error detection sequence.

The wireless environment is very noisy; since a corrupt frame has to be retransmitted. So, fragmentation is required for efficient data handling. A large network packet may be partitioned into smaller MAC frames before being transmitted. This process is known as **fragmentation.** The **MAC Protocol Data Unit (MPDU)** is created by the fragmentation technique to improve the chance of successful transmissions that are smaller than **MAC Service Data Unit** (MSDU). Each fragment is sent as an independent transmission and acknowledged separately. Once the station has occupied the medium, it will continue to send data fragments with an SIFS gap between the ACK reception and the start of the next fragment until all the fragments of a single MSDU have been sent or an ACK frame is not received. The source station maintains control of the channel throughout the transmission of the MSDU by waiting for only an SIFS period after receiving an ACK and transmitting the next fragment. If any fragment transmission fails, retransmission will occur after the back-off procedure. This is illustrated in Fig. 14.13.

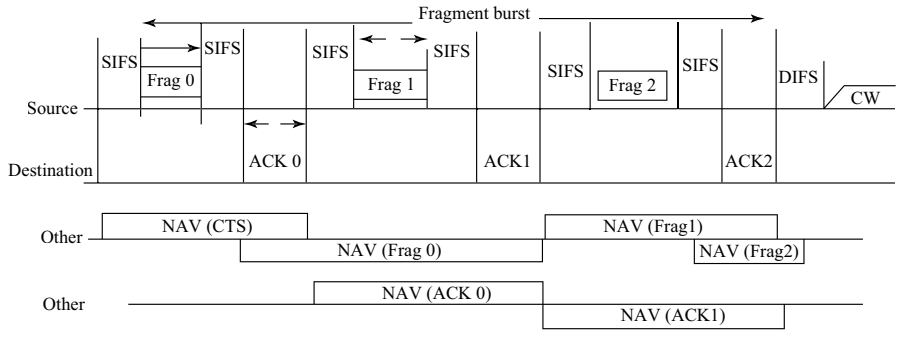


Fig. 14.13 *Fragment transmission procedure of MPDU © IEEE Copyright 1997*

14.9 IEEE 802.11 POINT COORDINATION FUNCTION (PCF)

The PCF provides un-contended access via arbitration by a point coordinator, which resides in the Access Point (AP). The DCF method provides best-effort service whereas the PCF guarantees time-bound service. Both methods may coexist, a contention period (CP) following a contention free period (CFP). PCF would be well suited for real-time traffic.

The PCF relies on the Point Coordinator (PC) to perform polling, enabling polled stations to transmit without contending for the channel. The AP performs the function of the PC within each BSS. The PC maintains a list of stations that need to be polled, and the PCF uses the Priority Interframe Space (PIFS) to capture and maintain the control of the medium. The period over which the PCF is operated is called the **Contention Free Period** (CFP). The PC tries to gain control of the medium at the beginning of the CFP after sensing the medium to be idle for a PIFS period. Once the PC gains control of the medium, it starts transmitting downlink packets to stations.

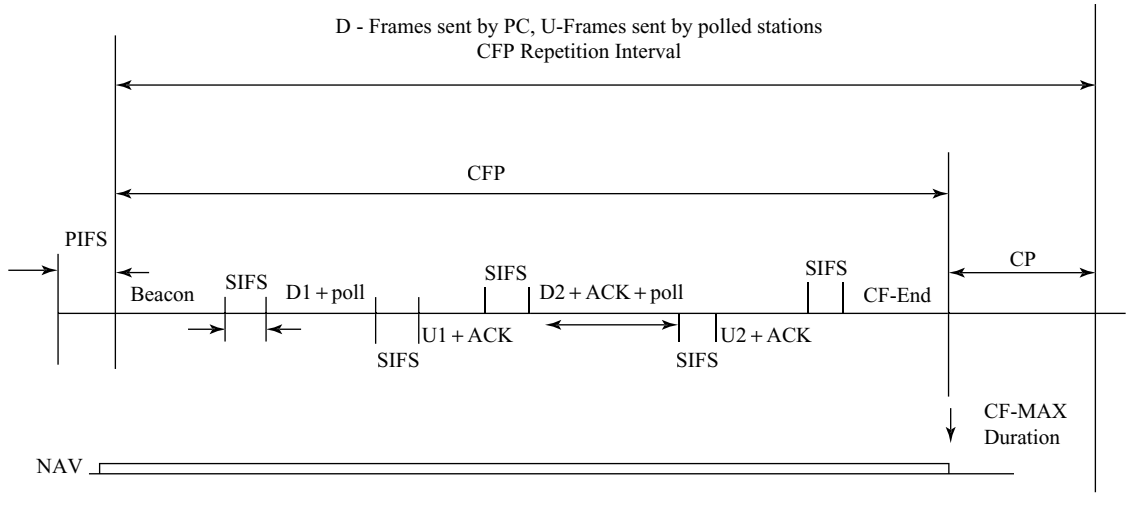


Fig. 14.14 *Timing diagram of successful transmission for PCF operation © IEEE Copyright 1997*

The PCF coexists with the DCF and logically sits on top of the DCF. The CFP repetition interval, which is the sum of CFP and CP, is used to determine the frequency of PCF operation. A beacon frame transmitted by an AP initiates the CFP repetition interval. Its function is to synchronize and maintain timing information. The CFP repetition interval is the integral multiple of beacon frames. The maximum size of

the CFP is determined by the manageable parameter CFP_MAX duration, which is the time required to transmit two maximum-size MPDUs including overhead, the initial beacon frame, and a CF-End frame. This is illustrated in Fig. 14.14. At the beginning of each CFP repetition interval, all stations in the BSS update their NAV to the maximum length of the CFP.

The PC can also send CF-poll (contention free-poll) frames to the stations that have requested CF services for their uplink traffic. Upon receiving a poll, the station transmits data after an SIFS period if it has uplink packets to send. Otherwise, it will respond with a NULL frame—a data frame without payload. For better utilization efficiency of the medium during CFP, both the CF-ACK and CF-poll frames can be piggybacked onto data frames [7].

14.10 IEEE 802.11 PHY LAYER

The IEEE 802.11 standards specify the MAC and PHY layers for fixed, portable and mobile stations within a local area. There are several specifications in the WLAN family: 802.11, 802.11b, 802.11a, 802.11g, 802.11e and 802.11i. The original standard supported three PHY IR, DSSS and FSSS as described in Section 14.1.

The 802.11b is the extension of the standard support DSSS in the 2.4 GHz band [7] with data rates of 1, 2, 5.5 and 11 Mbps. The original 802.11 DSSS standard specifies the 11-bit chipping, or Barker sequence, to encode all data sent over air. Each 11-chip sequence represents a single data bit of either 1 or 0, and is converted to a waveform, called a symbol, that can be sent over the air. These symbols are transmitted at 1 MSps (1 million symbols per second), using a sophisticated technique called **Binary Phase Shift Keying (BPSK)** [8] (Appendix A).

For a 2-Mbps data rate, a more sophisticated implementation called **Quadrature Phase Shift Keying (QPSK)** [9] (Appendix A) is used that doubles the data rate available in BPSK, via improved efficiency in the use of the radio bandwidth. For increased data rate in 802.11b (5.5 and 11 Mbps), **Complementary Code Keying** (CCK) is used [10].

CCK uses a set of 64 eight-bit unique code words—thus up to 6 bits can be represented by any code word (instead of the 1 bit represented by a Barker symbol). As a set, these code words have unique mathematical properties that allow them to be correctly distinguished from one another by a receiver, even in the presence of substantial noise and multi-path interference. The 5.5 Mbps rate uses CCK to encode 4 bits per carrier, while the 11 Mbps rate encodes 8 bits per carrier. Both speeds use QPSK as the modulation technique and signal at 1.375 MSps. QPSK uses four rotations (0, 90, 180 and 270 degrees) to encode 2 bits of information in the same space as BPSK encodes 1. Table 14.2 shows the different modulation schemes and data rates.

Table 14.2 *802.11 Modulations schemes and data rate specifications*

Data rate	Modulation schemes	Code length	Symbol rate	Bits/Symbol
1 Mbps	BPSK	11-Barker sequence	1 MSps	1
2 Mbps	QPSK	11-Barker sequence	1 MSps	2
5.5 Mbps	QPSK	8-CCK	1.375 MSps	4
11 Mbps	QPSK	8-CCK	1.375 MSps	8

A new high-speed physical layer, the IEEE 802.11a PHY [11], based on Orthogonal Frequency Division Multiplexing (OFDM) has been developed to extend the existing 802.11 standards. Unlike 802.11b, the 802.11a was designed to operate in the more recently allocated 5 GHz UNII (Unlicensed National Information Infrastructure) bands. Unlike the ISM band, which offers about 83 MHz in the 2.4 GHz spectrums, the IEEE 802.11a utilises almost four times that of the ISM band, because the UNII band offers 300 MHz of relatively free interference spectrum. The 802.11a can support 8 different PHY modes with data rates ranging from 6 Mbps to 54 Mbps. It is not backward compatible with the 802.11b based on DSSS technology. The FCC has divided the total of 300 MHz into three distinct 100 MHz domains, each with a

different regulated maximum transmission power as depicted in Table 14.3. The low, high band is suitable for intra-building transmission; whereas the high band is suitable for inter-building transmission.

Table 14.3 *Three bands of IEEE 802.11a*

Low band	Middle band	High band
5.15–5.25 GHz	5.25–5.35 GHz	5.725–5.825 GHz
TX-Power: 50 mW	250 mW	1.0 W

In the OFDM scheme, each sub-channel is about 300 KHz wide. The **OFDM** [12] works by breaking one high-speed data carrier (20 MHz wide and broken up into 52 subchannels, each 300 KHz wide) into several lower-speed subcarriers, which are then transmitted in parallel. Four channels are used for error correction whereas the rest are for data. BPSK is used to encode 125 Kbps, yielding a 6 Mbps data rate. Using QPSK, it is possible to encode up to 250 kbps per channel that can be combined to achieve a12 Mbps data rate. Again, by using 16-level Quadrature Amplitude Modulation (QAM) encoding 4 bits per Hertz, and achieving a data rate of 24 Mbps, the standard defines basic speeds of 6,12 and 24 Mbps, which every 802.11a-compliant product must support. Data rates of 54 Mbps are achieved by using 64-QAM (64 level QAM), which yields 8 bits/10 bits per cycle, and a total of up to 1.125 Mbps per 300 KHz channel.

The PHY layer specification for 802.11g [13] is almost the same as 802.11a, and uses OFDM technology but also supports DSSS in order to be backward compatible with 802.11b. It uses hybrid CCK and OFDM method. The header of a packet is transmitted in a single frequency (CCK) and the payload is transmitted in multiple frequencies (OFDM).

14.10.1 IEEE 802.11 PHY Sublayers

The PHY layer is divided into two sublayers—the physical-medium dependent sublayer (PMD) which carries out the modulation and the encoding, and the Physical Layer Convergence Protocol (PLCP). The PLCP presents a common interface for higher-level drivers to write to, and it provides carrier sense and CCA (Clear Channel Assessment), which is the signal the MAC layer needs to determine whether the medium is currently in use. The DSSS PHY layer packet format is given in Fig. 14.15.The complete packet is known as PPDU.

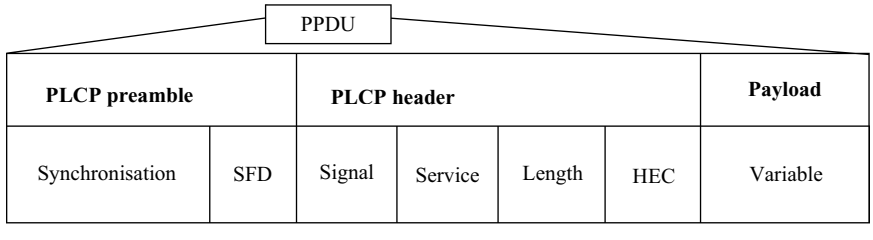


Fig. 14.15 *802.11b DSSS PHY layer packet format*

PLCP Preamble It is a 144-bit preamble and it forms from a SYNC field and SYNC Field Delimiter, used for synchronization to determine a radio channel and to establish CCA. This is PHY dependent, and includes the following fields:

SYNC It is 128-bit sequence of alternating zeros and ones, which is used by the PHY circuitry to select the appropriate antenna (if diversity is used), and to reach steady-state frequency offset correction and synchronization with the received packet timing.

SFD (SYNC Field Delimiter) It consists of the 16-bit binary pattern 1111001110100000 used to define frame timing and mark the start of every frame, and is called the SFD.

PLCP Header The header consists of 48 bits, is always transmitted at 1 Mbps and contains logical information used by the PHY Layer to decode the frame. It consists of the following fields:

Signal 8 bits containing only the rate information, encoded in 0.5 Mbps increments from 1 Mbit/s to 4.5 Mbit/s

Service 8 bits reserved

Length 16 bits and represents the number of bytes contained in the packet; used to correctly detect the end of packet

HFC (Header Error Check Field) 16-Bit CRC of the 48 bit header

14.11 802.11 SYSTEMS PERFORMANCE

In addition to the free space path loss model [14] (please see also in Chapter 3), multipath and shadow fading [14] have significant effects on the WLAN systems. The multipath effect is modeled by Rayleigh fading channel with a 50 ns delay spread, whereas shadow fading can be modeled as a lognormal distributed random variable with an 8 dB standard deviation. As 802.11b and 802.11g operate in the same frequency band, they have the same type of path loss model. But 802.11a operates at a higher frequency band and has a different path loss. The received signal strength in 802.11b and g has a 6.4 dB less path loss than the 802.11a devices [6].

The data rate of 802.11 systems is distance dependent. The transmission data rate is chosen based on the received signal strength. Based on the distance and propagation distance, a rate fallback algorithm is deigned to select the transmission rate. In general, a 10% error rate is often used as the fallback rate threshold. Table 14.4 shows the data rate vs coverage distance for 802.11 systems [6]. It gives an insight as to how data rate tradeoffs the coverage areas.

Table 14.4 *Data rate vs coverage area for different 802.11 systems*

802.11a		802.11b		802.11g	
Data rate Mbps	**Distance in feet**	**Data rate Mbps**	**Distance in feet**	**Data rate Mbps**	**Distance in feet**
54	24	11	115	54	42
48	35	5.5	180	48	47
36	83	2	225	36	65
24	88	–	–	24	85
12	170	–	–	12	135
6	225	–	–	6	180

Another important performance metric is the throughput. It is defined as the ratio of transmitted information bit to the transmission rate. Throughput depends on a number of factors—the transmission rate, protocol overhead, MAC efficiency, preamble and packet size. Again it depends on the contentions among multiple users, retransmission due to collisions and higher layer protocol efficiency such as (TCP/IP). The throughput is about 5 Mbps for an 11 Mbps data rate in 802.11b, and 30 Mbps for a 54 Mbps data rate in 802.11g and 802.11a.

14.12 SECURITY ISSUES: IEEE 802.11i

By default, IEEE 802.11 devices operate in an open system, where essentially any wireless client can associate with an AP without checking authenticity. The WLAN Link layer supports three authentication methods:

1. Static filtering based on MAC Address
2. Wired Equivalent Privacy (WEP)
3. 802.11i Standard

1. **Static filtering based on MAC address:** WLAN-APs drop traffic of all hosts except those of certain pre-configured network devices. Typically, filtering rules are specified using the layer-2 address (MAC address) of the network device.

   The disadvantage is that in a public environment with a dynamic user population, static configuration of all APs with a list of MAC addresses is infeasible.
2. **Wired equivalent privacy:** True authentication is possible with the use of the 802.11 options, known **as Wired Equivalent Privacy (WEP)**, where a shared key is configured into the AP and its wireless station. Only those devices with a valid shared key will be allowed to be associated to the AP. WLAN-APs verify whether the end host knows a shared secret key in the form of a 40 or 104 bit WEP key which is used for all network devices accessing the same AP. Shared-key authentication operates as follows:
   (a) A station requesting 802.11 services sends an authentication frame to another station.
   (b) Upon receiving the initial authentication frame, the station replies with an authentication frame containing 40/128 octets of the challenge text.
   (c) The requesting station copies the challenge text into an authentication frame, encrypts it with a shared key using the WEP service, and sends the frame to the responding station.
   (d) The receiving station decrypts the challenge text using the same-shared key and compares it to the challenge text sent earlier. If they match, the receiving station replies with an authentication acknowledgement. If not, the station sends a negative authentication notice. This method is illustrated in Fig. 14.16.

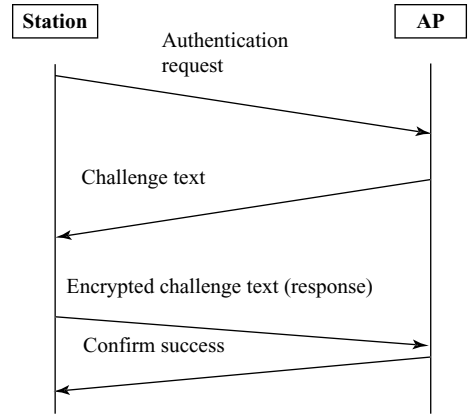


Fig. 14.16 *WEP authentication proces*

The main problem of WEP is that all users using the same AP share the same key. In a public environment it is very difficult to securely distribute and revoke the key. All users under an AP are authentic, and hence users can snoop on each other's traffic also, using the same key for encryption.

14.12.1 802.11i Standard

802.11i is a newer standard (yet to be finalised) for access control that allows dynamic per user per session authentication and encryption keys, and stronger packet encryption. It employs the IEEE 802.1x [15]-port access control standard that specifies the use of **Extensible Authentication Protocol (EAP)** over LAN (EAPOL) [16] between the mobile node and AP to perform the per user per session authentication procedure. EAP was designed to solve a major security problem, the assignment of an IP address after authentication in an IP network. EAP can be used over layer-2, over IP or any other higher layer; since it is an extension of point-to-point protocol (PPP). IEEE 802.11i also makes pre-authentication possible, while the current AP helps authenticating the new AP, thus supporting faster re-authentication and faster handover. When using 802.11i, the STA must first use Open System Authentication (OSA) with the AP. After OSA, the association takes place and the 802.1x process starts.

14.13 IEEE 802.11e: QoS ISSUES

In July 1999, the 802.11 Working Group initiated a study group to define a new standard referred to as the 802.11 within the standard known as 802.11e that provides Quality of Service (QoS) features in WLAN. The IEEE 802.11e [17] provides MAC enhancements to support local area networking with QoS requirements. There are two new modes of operation in the 802.11e—enhanced DCF and Hybrid Coordination Function (HCF). The QoS enhancements are available to the QoS enhanced stations (QSTAs) associated with a QoS Enhanced Access Point (QAP) in a QoS enabled network.

EDCF is based on the differentiating priorities at which the traffic is to be delivered. The access time that a station would need to sense the idle channel is variable. The 802.11e draft defines eight traffic categories for priority-based traffic. Each QSTA marks the packets for specific premium service requirements. The channel access parameter, i.e., inter-frame spacing (IFS) and contention window (CW_{min}, CW_{max}) differ from traffic to traffic. This mechanism, though providing QoS differentiation for traffic classes, does not guarantee QoS. Priority-based schemes may generate congestion than a reservation-based channel.

HCF is the enhanced version of the PCF that runs above the EDCF, and uses the PIFS to gain access over the medium. Instead of using a Point Coordinator (PC), there is the Hybrid Coordinator (HC), which is co-located with the QAP. The HC allocates the transmission opportunity to all QoS enabled stations—QSTA during the contention period and contention-free period. AIFS is larger than PIFS in order to get priority to HC over EDCF. Transmission begins during CP when the medium finds itself to be available under the EDCF rule after the AIFS plus backoff time, or when a station sends a special frame QoS CF-poll from the HC after PIFS idle period without any backoff. The HCF is used for applications such as voice and video that are needed for a periodic service from the HC.

14.14 SOME BASIC 802.11 SERVICES

In this section, some of the important 802.11 services like authentication, de-authentication, privacy, association, de-association, scanning and data transfer will be discussed.

1. **Scanning:** This is required to join a network, to initialise an ad-hoc network and finding an AP while roaming. Scanning may be active or passive. In passive scanning, a station listens to beacons from each channel that contain ESS-ID, BSS-ID and timestamp and saves those information. For active scanning, a station needs a probe in each channel and waits for a response that contains similar information as in passive scanning. The best AP is then selected to join the network. Active scanning is a faster process.
2. **Association:** Is needed between stations and the selected AP after scanning. The scanning and association process is shown in Fig. 14.17. The association service enables the establishment of wireless links between wireless stations and APs in infrastructure networks.

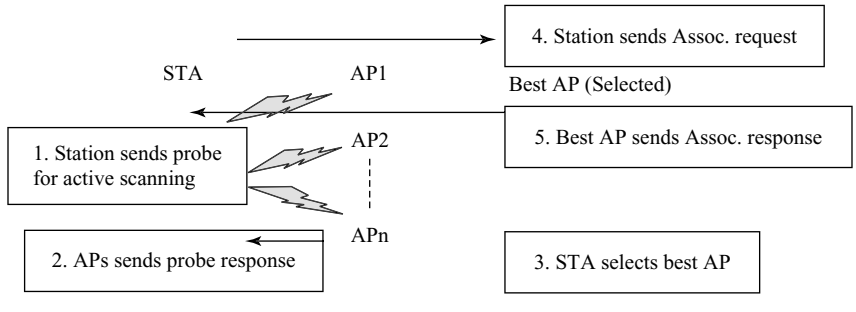


Fig. 14.17 *Association process*

3. **Disassociation:** The service that cancels the wireless links between wireless stations and APs in infrastructure networks is called *disassociation.*
4. **Authentication:** By authentication, WLAN services provide the client an identity to get access to the AP. The IEEE 802.11 defines open system and shared key authentication. The open system is a null authentication algorithm. With the help of WEP, the 802.11 can provide true authentication. A shared key is configured into the AP, and only those devices with a valid shared key will be allowed to access the AP. This authentication is only accessible if the WEP option is implemented. The secret shared key is to be delivered to participating stations via a secure channel independent of IEEE 802.11.

   To maintain **privacy**, the WEP option encrypts data before it is sent wirelessly, using a 40-bit encryption algorithm known as RC4 [5]. The same shared key used in authentication is used to encrypt or decrypt the data, allowing only wireless stations with the exact shared key to correctly decipher the data.

   The base station needed to deny client credential in case of fake users is based on the IP or MAC filters and performs **de-authentication**,
5. **Re-association:** The re-association service occurs in addition to association when a wireless station moves from one BSS to another. Two adjoining BSSs form an ESS if they are defined by a common ESS-ID, providing a wireless station with the capability to roam from one area to another. Although re-association is specified in 802.11, the mechanism that allows AP-to-AP coordination to handle roaming is not specified.
6. **Power management:** This is required to save battery power in wireless application. The 802.11 standard directly addresses the issue of power saving and defines a mechanism for the wireless station to go into sleep mode for a long period of time without losing any information. The AP maintains updated information of the current stations working in power-saving mode. Until a station requests for the packets addressed to these stations by sending a polling request, or changes their operational mode, the packets are buffered at the AP.

14.15 ROAMING, HANDOVER AND MOBILITY MANAGEMENT FOR WLAN

The 802.11 MAC layer is responsible for how a mobile station associates with an access point. The standard includes mechanisms to allow a station to roam among multiple APs that can be operating on the same or separate channels.

To synchronize other STAs within the BSS, an AP transmits a beacon frame where time stamp is present; this prevents some drift of the STA synchronization. The beacon frame can be deferred only if a station is transmitting data. A roaming station searches the strength of the signal for the best AP it wants to be connected to. If it finds weak connection, then it establishes a new association with the new AP through the scanning process.

Due to mobility of a station, the STA can change the BSS where it is to be connected using the active/passive scanning and re-association service. If the search is successful, the STA re-associates with the new AP by sending an association request and getting a response from the AP. If re-association fails, STA searches for another BSS. For a roaming STA associated to new AP, the latter informs the DS (Distribution System), and normally an old AP is notified through the DS.

14.15.1 Handover and Mobility Management

Handover is a part of mobility management that occurs due to the mobility of a wireless station. Mobility management supports roaming of the station allowing uninterrupted services by getting connections to new access points. Handover management focuses on the exchange of control signals between old and new access points during active data transmission.

Mobility affects all protocol stacks from physical, data link, network to transport and application layers. The goal of mobility management is to ensure some of the key functionalities like fast handover with low handoff delay, seamless connection with less or no packet loss, less signaling overhead for better scalability, efficient routing, quality of service maintenance and providing security.

When a mobile station moves from one point of attachment to another, it requires re-association service to a new AP and de-association from the old AP. Other network elements involved in providing the session should also be informed to continue seamless services.

14.15.2 IEEE 802.11 Handover Scenarios

Figure 14.18 illustrates the different handover scenarios in IEEE 802.11.

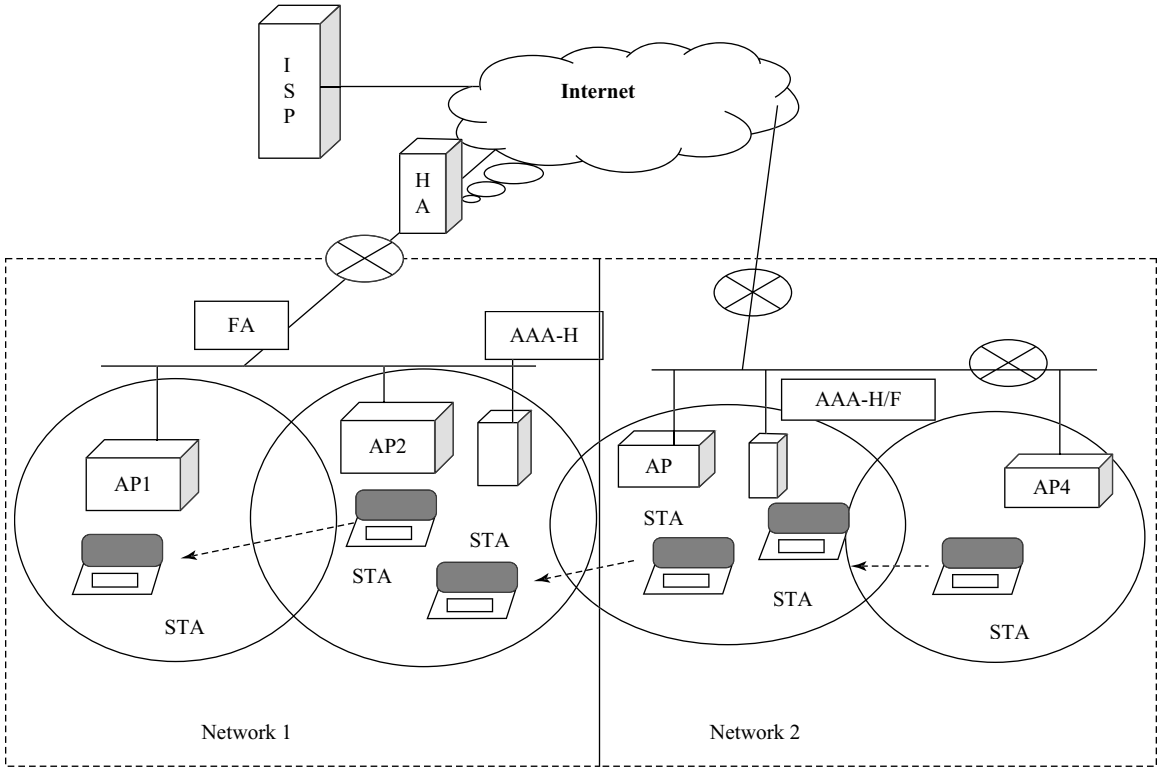


Fig. 14.18 *Handover scenarios for 802.11*

1. Mobility within a Basic Service Set (BSS), Local Mobility under the coverage of one AP
2. Mobility from one AP to another AP under the coverage of Extended Service Set (ESS) and intra ESS mobility
3. Mobility of the station from one BSS under an ESS to another BSS under different ESS, called inter-ESS mobility.

The first two are very common to WLAN and support link layer (Layer 2) handover. The third one needs a network layer handover.

Consider that all access points are on a single subnet, so the roaming only happens at Layer 2 and the roaming device keeps the same IP address.

When a WLAN STA moves from the range of one access point to another in the same subnet (such as **STA1 moves from AP2 to AP1** as shown in Fig. 14.18), it needs to find the best AP, decide when

to roam onto it, associate with it and do any **authentication** required, as per the security policies. Then the wired network (distribution system) has to learn the location of the STA, so that data can be sent to it. The scanning and decision-making part of the roaming process allows the STA to find a new AP on an appropriate channel frequency within 802.11b, either with the same or a different channel as the user moves. An STA searches an AP with a better communication link (received signal strength), and when this happens, the STA must associate with the new AP. The searching station can initiate an active mode by sending a probe request message referencing the target set of interconnected APs. Each AP will respond with a probe response message, which serves as the solicited beacon.

Thus, the specific actions which occur as a user roams from one AP to another are summarized as follows:

1. The station sends a **re-association** request to a new AP.
2. If the re-association response is successful then station has roamed to the new AP otherwise, the station **scans** for another AP.
3. If the AP accepts a re-association request after the required authentication, the AP, indicates **re-association to the distribution system**, the DS information is updated, and the old AP is notified through the DS.

Re-association is required when the roaming STA physically moves to the new AP, or there are changes in the radio characteristics in the building, or there is a congested channel due to high traffic, triggering handover state. This process of dynamically associating and re-associating with APs allows a customer to set up WLANs with very broad coverage by creating a series of overlapping 802.11b cells throughout a building or across a campus as shown in Fig. 14.18.

The APs deploy an **Inter Access Point Protocol (IAPP)** [18] to inform each other about station handovers. The IAPP is required for the rest of the network to be made aware that the STA has moved. This calls for **AP-to-AP** communication, which was never supported in the original 802.11 specifications, but is supported in the **802.11f** standard. The IEEE 802.11f standard IAPP is a communication protocol used for AP-to-AP communication. It comprises the part of communication systems with APs, STAs, backbone network and the **RADIUS** [19] **(Remote Authentication Dial-In User Service)** infrastructure. The RADIUS server provides the following two functions:

1. It maps the AP's ID to it's IP address.
2. It distributes keys to the APs to allow the encryption of communications between the APs.

IAPP calls for the new servicing AP to send out two packets onto the wired LAN. One of these is the source address of the STA (usually broadcast, however some implementations still use unicast to the previous AP or a multicast) and is used by intervening switches to **update their MAC address tables with the roamed STA** to the new location. The second is an IAPP ADD-notify packet from the new AP to an IAPP multicast address that all APs subscribe to, and which contains the MAC address of the station it has just associated. All APs will receive this packet, and the one that had been associated with that station will use the sequence number included to determine that this is newer information and **remove** the stale association from its internal table (de-association). IAPP provides for the sharing of information between APs and provides proactive caching that is a method to support fast roaming by caching the context of an STA to the APs to which the STA may roam. The next APs are identified dynamically by learning the identities of the neighbouring APs.

To complete all these processes, some time is required which is known as the *handover delay* for inter-AP movement. The components for handover delays are listed as

1. Detection of handover requirement
2. Active or passive scan
3. Re-authentication (for 802.1x supplicant, re-authentication with the RADIUS server)
4. Re-association depending on handover scenario, layer-3 and higher layer handoff

For a single channel, there is no detection and passive scan delay. For a multi-channel, active scanning causes delay. It is dependent on the traffic condition. Re-authentication and re-association processes contribute significant time for handover delay. For inter domain handover, layer-3 handover due to Mobile IP [20] operation will be added. For further details of handover in 802.11 systems, the reader may go through [21].

14.16 WLAN APPLICATIONS

The target market for WLAN will be both in private and public domain in the form of hot-spot regions for low-priced, high data-rate services. Public hot spots refer to WLAN places in public areas like libraries, cafes, airports or railway stations. Private hotspots may be in hospitals, shopping malls, gas stations or universities. WLAN has brought about many possibilities in inventory management for enterprise environments.

The emergence of voiceover IP technology in wire-line networking presents an opportunity for combining packetized voice and data over one convergent network infrastructure. Wireless operators are also looking for WLAN hot spots, both for voice and data services. Voice-over WLAN makes wireless voice application less expensive, and easier to install and maintain. The mobile operators need to play a major role in pushing the WLAN market to public services in hotspots. The interoperability and service agreement among Wireless Internet Service Providers (WISP) will allow users to move from one WLAN network to other. Before that we need a solution that seamlessly provides handover from one access point to another efficiently along with a solution to the security problem.

Summary

The fundamental operations of WLAN and a brief overview of the WLAN family are provided in this chapter. The growth in wireless and its benefits has brought forward the use of WLAN in the world, both in private and public places. The IEEE 802.11b based WLAN is already becoming popular, while 802.11a and 802.11g are also being used. In the near future, QoS based WLANs are expected to enter the market. With the increasing demand in mobile wireless communication and development and standardization beyond the 3G (B3G) system, WLAN will play a major role in integrating with 3G cellular systems to provide the complementary benefits of low cost and high data-rate services.

References

[1] *IEEE 802.11, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, November 1997.

[2] *IEEE 802.11 WG, Part II: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications*, IEEE Standard, Aug. 1999.

[3] *IEEE 802.11 WG, Part II: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, High–speed Physical Layer in the 5 GHz Band, Supplement to IEEE 802.11 Standard*, Sep. 1999.

[4] Forouzan, B.A, *Computer Communication and Networking*, McGraw-Hill, 2004.

[5] Nedeltchev, P, *WLANs and the 802.11 Standard*, White Paper, March 2001.

[6] Chuah, M.C, and Q Zhang, *Design and Performance of 3G Wireless Networks and Wireless LANs*, Springer, 2006.

[7] *ISO/IEC 8802-11, ANSI/IEEE Standard 802.11, First Edition 1999-00-00, Information Technology–Telecommunications and Information exchange between systems–Local and metropolitan area networks–Specific requirements-Part II: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, DSSs.*

[8] http://www.physics.udel.edu/wwwusers/watson/student_projects/scen167/thosguys/psk.html

[9] http://www.ee.byu.edu/ee/class/ee444/simulink/oqpsk/oqpsk.html.

[10] LaMaire, R.O. et al, *Wireless LANs and Mobile Networking: Standard and Future Directions*, IEEE Communication Magazine, Vol. 34, No. 11. pp. 86–94, Aug 1996.
[11] *IEEE Std 802.11a-1999, Supplement to IEEE Standard for Information technology–Telecommunications and information exchange between systems–Local and Metropolitan area networks–Specific requirements–Part II: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High-speed Physical Layer in the 5 GHz band*.
[12] Van Nee, R.D.J, and R. Prasad, *OFDM for Wireless Multimedia Communications*, Norwood, MA, Artech House, 2000.
[13] *IEEE 802.11g, IEEE 802.11: Further High Data Rates Extension in 2.4 GHz band, D11.2,* April 2003.
[14] Rappaport. T.S, *Wireless Communications, Principles and Practice*, Prentice Hall, PTR, 1996.
[15] *IEEE Standard 802.1x-2001, IEEE Standard for Local and Metropolitan area networks–Port based Network Access Control,* 14, June 2001.
[16] Blunk, L, J. Vollbrecht and B. Aboda, *Extensible Authentication Protocol (EAP)*, Oct. 2002, IETF pp-pext working group draft, draft-ietf-pppext-rfc2284bis-07.txt
[17] *IEEE 802.11 WG, Part II: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Medium Access Control Enhancements for Quality of Service, D4.4*, June 2003.
[18] http://www.ieee802.org/11/.
[19] Rigney, C, A. Rubens, W. Simpson, and S. Willens, *Remote Authentication Dial In User Service (RADIUS), RFC 2138*, Jan 1997.
[20] Perkins, C, *Mobile IP, IEEE Comm.*, Vol. 35, No. 5, 1997, pp. 84–911.
[21] Prasad, A.R, and N.R. Prasad, *802.11 WLANs and IP Networking, Security, QoS and Mobility,* Artech House, 2005.

Questions for Self-Test

14.1 CSMA/CD is suitable for wired media but not for wireless transmission.
a. False b. True

14.2 CSMA/CA is a suitable protocol for wireless LAN.
a. True b. False

14.3 A random back-off time is used following a busy medium condition for CSMA/CA.
a. False b. True

14.4 The network allocation vector maintains a prediction of future traffic on the medium.
a. False b. True

14.5 To reduce transmission error probability, RTS/CTS method is useful.
a. False b. True

14.6 CSMA/CD allows collision, CSMA/CA avoids collision, throughput reduces with 30–40% increase in collision for
a. CSMA/CA b. CSMA/CD

14.7 RTS/CTS frames are sent prior to actual data transmission for medium reservation.
a. True b. False

14.8 _____ created by fragmentation improves the chance of successful transmission.

14.9 The size of RTS is _____ octets and CTS is _____ octets.

14.10 The MAC layer of 802.11 standard for CSMA/CA defines two access methods: _____ and _____.

14.11 The value for SIFS is _____, PIFS is _____ and DIFS is _____.

14.12 WLAN uses the unlicensed ISM frequency band named as _____.

14.13 The ISM band has three frequency ranges as _____ , _____ , _____.

14.14 The most widely used WLAN standard is _____ work at the frequency band _____.

14.15 The theoretical data rates of IEEE 802.11b and IEEE 802.11g are _____ , and _____ respectively.

14.16 IEEE 802.11 standards define the _____ and _____ layers.

14.17 IEEE 802.11 physical layer defines three physical techniques for wireless LAN. They are, _____ , _____ and _____.

14.18 IEEE 802.11 standard provides _____ hops pattern or frequency shifts to choose from in the 2.4 GHz ISM band.

14.19 In directsequence spread spectrum, signals are transmitted at a given _____ with a specific _____ so that the receiver can identify the appropriate transmitter.

14.20 In DSSS, the bandwidth allocated to each station is much _____ than the required bandwidth.

14.21 The 802.11 standard uses CSMA/ _____ , whereas standard Ethernet uses _____.

14.22 Specify the unlicensed frequency bands for WLAN available in 2 to 6 GHz ranges worldwide.

14.23 What are the different modulations used in WLAN? What are the limits of data rate to achieve?

14.24 What is the channel bandwidth in the 2.4 GHz band for 802.11 standard?

14.25 Describe the three different techniques for infrared transmission in WLAN.

14.26 What types of WLAN are used at your workplace? Identify the frequency band, channel bandwidth and modulation techniques used for that.

14.27 Describe the operation of DCF. In which situation is PCF used?

14.28 If there is a hidden station problem and if two stations cannot hear each other during transmission, what will happen? How can this situation be eliminated?

14.29 Describe the basic WLAN architecture and it's components.

14.30 Why is CSMA/CD not suitable for wireless networks?

14.31 Contention window influences performance for back-off algorithm–explain.

14.32 What are the services that the IEEE 802.11 standard can offer?

14.33 Describe the MAC layer architecture and functionality for IEEE 802.11.

14.34 How does use of RTS and CTS help in avoiding collisions.

14.35 Describe the different inter-frame spacing defined in 802.11 and also define the significance of those frames.

14.36 What are the functions of DCF and PCF?

14.37 With a schematic diagram describe the timing of successful transmission for PCF operation.

14.38 What are the different security measures that can be taken in WLAN?

14.39 What is the special purpose of IEEE 802.11i standards?

14.40 Security is a big threat for WLAN technology. How can this threat be handled to increase reliability?

14.41 What are the different types of handoff occuring due to the mobility of the roaming users within a WLAN environment?

14.42 When is Inter Access Point Protocol used? How does this protocol help in handover?

14.43 Discuss the mobility management procedure in IEEE 802.11 standard.

14.44 What is the Inter Access Point Protocol? Why is it needed?

14.45 Discuss some applications of WLAN.

Cellular and WLAN Integration: Heterogeneous Network Architecture, Step towards 4G Networks

15

Introduction

Currently, various wireless technologies and networks exist that fulfill different needs and requirement of mobile users with respect to Quality of Service (QoS), radio coverage, multimedia service and data rate. Wireless LAN is a satisfactory solution for high-data-rate hot-spot access. Traditional and next generation cellular networks provide medium-data-rate wide-coverage service. Hence, integrating the complementary systems enables the best connection of mobile terminals anytime and anywhere. The discrepancy of access technologies and network architecture demand that the common infrastructure to integrate diverse wireless networks.

Nomadic users need to have ubiquitous access to remote information storage and computing services. As an evolutionary step toward 4G mobile communications, mobility in heterogeneous IP networks with GPRS, UMTS and IEEE 802.11 WLAN systems is seen as one of the central issues in the realization of 4G communication network and systems.

Cellular (GPRS/3G) and Wireless Local Area Networks (WLAN) dominate wireless communication currently. Though, the cellular system is the main wireless media primarily for voice communication, it is recently being enhanced for data services. Contemporarily, the use of WLAN has increased in the hot-spot regions due to its high data rate and relatively low installation cost. So, integration of the two complementary systems is the focus for low-cost high-data-rate multimedia communications.

In 4G networks, to suffice the needs of users with relatively low mobility but yet to provide a substantial amount of bandwidth in hot spots, it becomes essential to integrate WLAN 802.11 systems with currently existing UMTS-GPRS systems. The integrated network can handle wider mobility of the mobile nodes (MN). The 3G cellular networks, e.g., UMTS are designed to provide users with voice and data services. The total cell capacity limits the per user data rate. Many a times high-speed requirements are clustered in small pockets. These clusters are termed **hot spots**. Network operators would like to employ efficient solutions, which can be easily integrated with their existing UMTS-based infrastructure. WLANs, therefore, offer an attractive solution. 3G cellular accesses based on code division multiple access (CDMA), either wideband CDMA or cdma2000, can be used to satisfy users who have a larger need for mobility, while 802.11 systems can be used to support users with much lesser coverage-area requirements. It is in light of this that the next wave of technological advance is already under consideration, i.e., 4G. Several international projects are already underway to ensure that services are suited to the characteristics of different delivery mechanisms, from cable networks to GPRS networks; and that the services can be delivered using the most appropriate of the range of networks available. Both the 3GPP and 3GPP2 experts are developing specifications for inter-working architectures that enable users to access their 2G and 3G data services from WLAN hot spots. In addition, the dynamic allocation of available spectrum between different wireless networks is also under investigation. The challenge is to design such

a transport infrastructure which will be able to take full advantage of IP-based technologies achieving desired mobility between the various access techniques and at the same time provide the necessary capabilities in terms of QoS, robustness and manageability. The goals at the present stage regarding the development of mobile standards remain common (3GPP and 3GPP2) and include IP-based multimedia services, IP-based transport and the integration of IETF protocols for functions such as wide-area mobility support (MIP-Mobile IP), signaling (SIP-Session Initiation Protocol) and authentication, authorization and Accounting (AAA). It is popular to call any network that satisfies these criteria as an all-IP network. In the near future, mobile access to Internet will be the combination of heterogeneous networks. The most attractive solution will be the cellular-WLAN integration that needs to cater to vertical handoff for uninterrupted services, and the terminal must support dual interfaces. In this chapter, we shall provide the basic integration techniques and their pros and cons. Mobility management will become a difficult task for heterogeneous networks. This chapter will also discuss the basics of handoff management for integration architecture.

15.1 WHY INTEGRATION?

1. Operators would like to take advantage of the growing deployment of WLAN hot spots and their complementary nature with 3G networks.
2. For data services, mobile operators are interested in roaming between WLAN and 3G networks enabled with dual-band, dual-mode handsets.

The integration aspects of interest to operators are:

1. Access-independent services
2. Low-cost infrastructure for high capacity and coverage
3. Better utilization of spectrum resources
4. Standardised inter-working functionality
5. Promotion of multimode terminals
6. Flexible integration of new access networks at the IP level

The complementary nature of the WLAN and 3G technologies can give cost effective business.

1. Integration of the best of two can provide
   (a) Efficient use of radio spectrum
   (b) Better coverage
   (c) Integrated voice services
   (d) Seamless wireless data services and common billing
2. Possible applications

Wireless multimedia and other high-data-rate services for a large population

15.1.1 Benefits of Integration

1. **For 3G service provider:** Economically offload data traffic to WLAN in areas with high user density
2. **For WLAN service provider:** Bring a larger user base from partner 3G networks
3. **For end user:** Enhanced performance through greater coverage, higher data rate and lower overall cost.

By linking WLAN to cellular networks, mobile operators can provide significant customer value in areas like

(a) *Authentication and billing relationship:* The customer will prefer a single billing system and service support
(b) Secure communication environment and trusted relationship

(c) Seamless roaming and handoff between WLAN and GPRS/3G networks
(d) Access to a wide range of applications and services (MMS-Multimedia services)

15.2 INTERNETWORKING NETWORK ARCHITECTURE

The most pertinent question that comes in mind for integration architecture is 'where to integrate and how to integrate'? The focus must be on keeping the technologies intact or to bring very minor changes to the existing systems. The implementation involves incorporating new entities and protocol stacks that operate at the network layer or higher layers in order to enable internetworking without affecting the physical and data link layer of the existing systems.

In Cellular–WLAN integration, five basic implications are terminal equipment, authentication procedures, service and its QoS parameter, mobility management and uniform billing. The air interface card and protocol stack of the **Mobile Subscriber (MS)** of cellular packet networks and those of WLAN access networks are different from each other. This implies that for vertical roaming, the MS must be supported with the WLAN air interface card. The GPRS/UMTS supports QoS for four well-defined service classes—interactive, voice, streaming and best-effort traffic. In contrast, the QoS support for 802.11 is still under discussion in IEEE 802.11WG. CSMA/CA is suitable for best-effort traffic and PCF (Point Coordination Function) supports isochronous traffic. The issues which are rigorously kept in mind for designing a dual terminal equipment, are related to retaining of present technology to a maximum possible extent, less computation and complexity at terminal equipment and providing all inter working, mapping/conversion functionality at access networks.

15.2.1 Overview of GPRS/UMTS Networks

4G networks can be modeled as a basic UMTS/GPRS domain within which there are islands of 802.11 WLANs to serve the hot spots created by user requirements in particular locations. The basic UMTS network as shown in Fig. 15.1, provides packet-switched (PS) data services and circuit switched (CS) voice service.

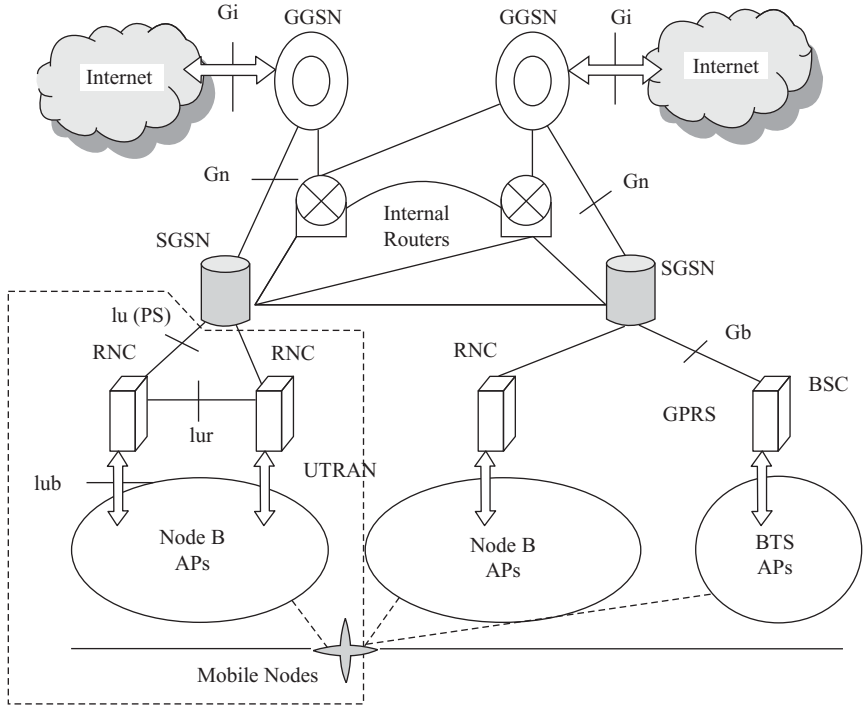


Fig. 15.1 *UMTS network*

The GPRS network is integrated into UMTS to get the PS service. In UMTS, RNCs (Radio Network Controller) and Node Bs constitute the Radio Access Network (RAN) called UMTS RNS.

Each Node B constitutes of a cluster of base stations or APs (Access Points), and a number of Node Bs may be connected to a single RNC. The packet core network (CN) comprises of SGSN (Serving GPRS Support Node) and GGSNs (Gateway GPRS Support Node). In RAN, the RNC receives downlink packets from SGSN and converts them to radio frames before sending them over to the Node Bs. In the uplink direction, the RNS receives the radio frames from the Node Bs and converts them to IP packets and dispatches them to the SGSN. The RNC manages the radio resources at the Node Bs and also is responsible for the maintenance of RABs (Radio Access Bearer) through them. The SGSN also maintains the mobility context of the UE (User Equipment)/MN and controls RAB set-up through RNCs. Two GTP (GPRS Tunneling Protocol) sessions are created between SGSN–GGSN and SGSN–RNC for passage of IP packets. When multiple APs serving an MN have different controlling RNCs then one of the latter acts as the serving-RNC for that host. It is this serving-RNC that is responsible for frame selection among the multiple received copies of the same transport block.

The MN attaches to the CN by the GPRS-attach operation. This creates two GTP sessions between SGSN–GGSN and SGSN–RNC. Upon GPRS attachment, a mapping is created at the RNC between the host ID and GTP session, and RNC and SGSN. Another mapping is created at the GGSN between host's network IP address and SGSN. The SGSN handles inter-RNC mobility and the GGSN handles inter-SGSN mobility.

15.2.2 IEEE 802.11 Overview

The IEEE 802.11 standard accounts for ad-hoc and infrastructure modes. In infrastructure mode, an AP performs point coordinate function by forming a cell in it's coverage area, called *service set*. An MN can associate with only one AP at a time. All the MNs connected to an AP communicate either via the AP or are directly coordinated by the AP. Roaming across APs is supported in MAC. APs generate periodic beacons that advertise the network ID (SSID) and the cell ID (BSSID, AP MAC address). At power up, the MN sends an ASSOCIATE request to the AP. When the MN moves to a different cell, it receives a beacon with the same network ID but a different cell ID. It associates with the new AP by sending the MAC address of the old AP in the ASSOCIATE frame. Figure 15.2 shows the basic set of WLAN systems.

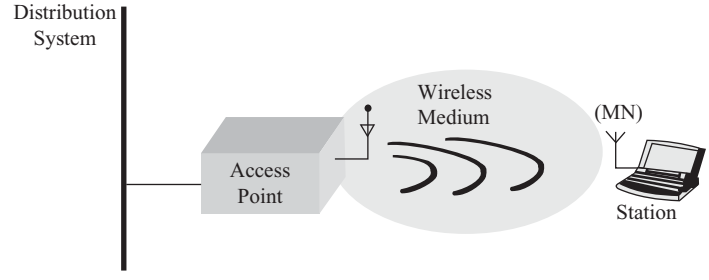


Fig. 15.2 *Wireless LAN systems*

The main components for WLAN systems are the following:

1. **Stations:** Typically battery operated laptops or handheld computers.
2. **Access points:** Generally fixed, performs wireless to wired bridging function.
3. **Wireless medium:** RF and IR (infrared).
4. **Distribution system:** Backbone networks used to integrate APs.

15.2.3 Complementary Features of Cellular and WLAN

3G networks will provide at best a 2 Mbps data rate at the user level which is much below the required speed for business multimedia service, while. WLAN can provide very high-speed data communication at the user level. Therefore, WLAN can be used as a complementary access network to 3G cellular networks for providing high-speed data communication at the user level. WLAN supports very low mobility; with

higher mobility, data rate will be much slower. Therefore, WLAN can best be used in hot spots on behalf of 3G access networks to provide a very high-speed data rate to 3G users. Large area coverage and higher mobility support will restrict the highest bit rate offered by 3G. Some of the key issues for 3G and WLAN are compared in Table 15.1.

Table 15.1 *3G vs WLAN*

Issues	UMTS/GPRS systems	WLAN system
Signaling	Out-band signaling	In-band signaling
Coverage	2-10 miles	100 ft
Date rate	2 Mbps static users, 144 kbps mobile	11 Mbps (802.11b) and 54 Mbps (802.11a)
Connection-to-core network	MN is connected to the core network via BSS by circuit switched and virtual circuit-switched method.	MN is connected to the core network via access point (AP) and it can access a channel only when the whole channel bandwidth is free for a certain time (802.11).
Mobility management	Mobility management entity resides in layer three sublayers (below the IP layer).	In WLAN, at the link layer, abstract mobility management, association and de-association are implemented. The mobility management entity (HA/FA) has been implemented in IP layer.
Protocol stack	Control and user protocol stacks are different.	Control packets and user data are processed through the same protocol stacks.
Flow control	Flow control exists for downlink BSSGP PDUs.	AP does not provide flow control for downlink traffic
QoS	Supported in access networks (bit rate, precedence, transmission mode, etc). The GPRS/UMTS supports QoS for four well-defined service classes—interactive, voice, stream and best-effort traffic.	No QoS provision in access networks. QoS support for 802.11 is still under discussion in IEEE 802.11WG.

15.2.4 Suitable Point of Integration

There is no single integration architecture that is good for all integration scenarios. Connecting WLAN at BSS/RNC requires major revision of complex radio procedures implemented at RNC/BSS because the two radio interfaces are totally different. WLAN can be emulated as a BSS/RNC by connecting it through a Gb interface at the SGSN as shown in Fig. 15.3. Although in this method, mobility management can be performed by GPRS/UMTS networks, the high-speed data rate of WLAN is injected into GPRS/UMTS core networks that may cause bottlenecks in the networks. All traffic will flow through SGSN and GGSN even if the destination is in the same WLAN causing unavoidable excessive overhead. Alternatively, the WLAN can be connected at the GGSN that seems to simplify the handover from GPRS/UMTS to WLAN because the GGSN only maintains session contexts for packet-switched connections. But in this case, during handover from WLAN to UMTS, the SGSN needs to recreate the mobility state and re-establish the session (PDP) and BSC/RAB contexts; these are the information that the GGSN does not have, hence the handover would be slow. The most promising solution is that the WLAN is to be connected to a separate IP-based network,

which in turn is connected to GPRS/UMTS networks through a Gi interface. This method is used for IP based loose coupling integration method, which requires least modification in either terminal equipment or in networks. In the following sections, the mode of integration will be discussed thoroughly.

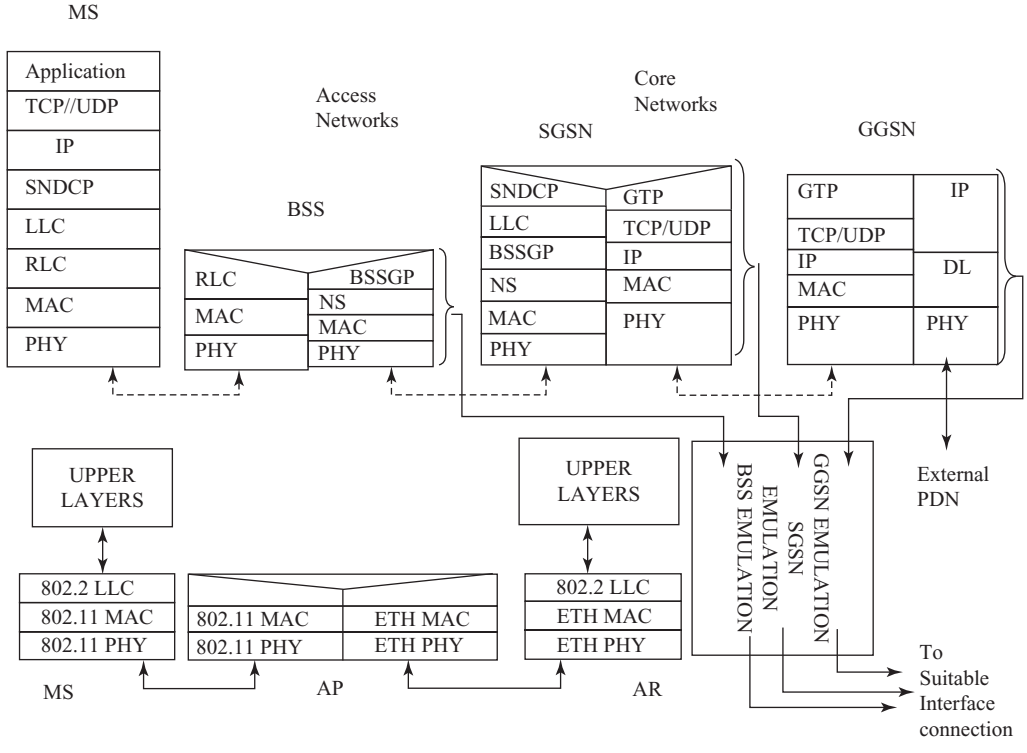


Fig. 15.3 *Different point of integration of WLAN/GPRS networks*

15.2.5 Integration Architecture

A detailed overview describing the internetworking between 802.11 WLAN systems and UMTS-based systems has been provided in [1]. Development of criteria to select the integration points between two such networks with different radio frameworks is a major concept which was explored in this literature. There are two main broad categories of integration architecture.

1. Tightly Coupled Integration Architecture (**TCIA**)
2. Loosely Coupled Integration Architecture (**LCIA**)

1. Tight-Coupling Integration: In tight coupling, the WLAN network acts as another 3G-access network. WLAN traffic is directly injected into the 3G cores. The set-up of the entire network and the design of SGSN and GGSN have to be modified to handle the increased load. There are three possible ways of tight-coupling integration as shown in Fig. 15.4.

(a) Tight coupling at the GGSN level: WLAN, instead of using its own packet-data gateway (PDGW), uses the 3G network's GGSN as its gateway to PDN. The WLAN network is seen as a RAN (Radio Access Network) by the 3G networks.

(b) Tight coupling at the SGSN management level: A group of WLAN APs, called ESS (Extended Service Set), forms a routing area and is seen by the 3G networks as a UTRAN (UMTS Radio Access Network)

Routing Area (RA). The ESS is connected to an SGSN via a GPRS inter-working function (GIF) using Gb or Iu-PS interface. The GIF makes the ESS look like a typical RA composed of one cell.

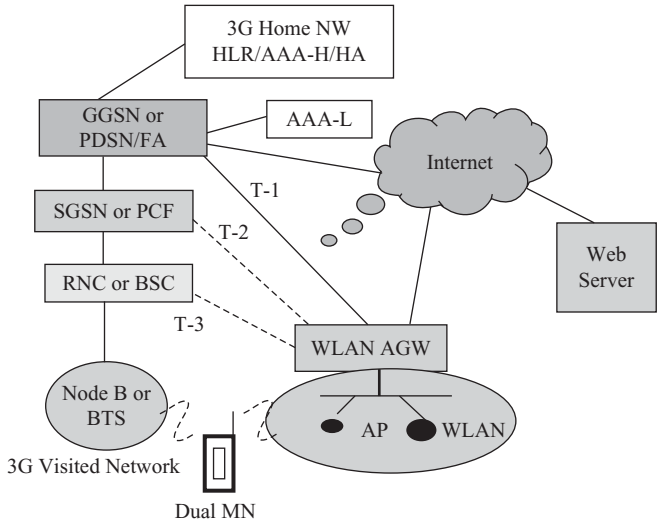


Fig. 15.4 *Tight-coupling integration architecture*

The dual stack user equipment (UE) scans for a beacon signal to see if the WLAN belongs to the wireless operator. If it does, the UE associates itself to the WLAN, then does an (Routing Area) update to the 3G network. The exchange of packets is via GIF using MAC addresses and forms the 3G CN perspectives. Handover between 3G-WLAN is considered as a handover between two cells.

(c) *Very tight coupling at the RNC cell management level:* A new Iu RNC-WLAN interface is introduced such that WLAN is seen as a cell at the RNC level. This requires specific NICs (Network Interface Cards) in the UEs.

Now the question is that which is the best point of integration? Let us discuss the suitable architecture of integration.

One of the requirements of the architecture specification can be that the user in a WLAN network can use the 802.11 accesses for UMTS PS (Packet Switched) service and the UMTS RNS for the UMTS CS (Circuit Switched) service. One can maintain access to both connections in parallel but cannot access the UTRAN for PS service, while the user in the UMTS network can use UMTS RNS for both PS and CS services. Once the MN moves into the 802.11 cells, the PS connection via the UMTS domain is dismantled and is re-established through the WLAN network. Inside the WLAN, the mobile node (MN) can use the UMTS RNS for CS service, and thus it becomes mandatory for the MN to possess two distinct interfaces, one for the UMTS-GPRS and one for the 802.11 networks. We must understand that the radio specifications for both 802.11 and UMTS-GPRS are different, and it would require considerable effort to build inter-compatible systems. Hence, the integration point can certainly not be the RNC/3G-AR. A better logical choice would be to treat the SGSN at the next highest level as the integration point since it is the next most closest point to the wireless stack (see Fig. 15.4). Also the GGSN has not been chosen as the integration point because at the time of handover to UMTS, the SGSN needs to recreate a mobility state and acquire session PDP and RAB context. These are information that the GGSN does not have, and hence it would be inefficient to choose it as the integration point. The MN must have two interfaces, and therefore there has to be a layer of abstraction, say an IP handler layer between the UMTS and 802.11 device drivers, and the normal IP layers to extract device-specific context information for both the interfaces.

Disadvantages of TCIA TCIA produces delay in packet delivery when the distance between the home network and the foreign network is very long, and hence is not suitable for real time communication since it increases the overall traffic of the GPRS core network.

Since the client device should be configured with both WLAN card (NIC) and SIM card there is a security problem. Also, the UMTS core network is directly exposed to WLAN networks.

2. **Loose coupling integration:** Figure 15.5 shows the loose-coupling integration architecture. In loose coupling, there is a common customer database and an authentication procedure. In this mode, the operator will still be able to utilize the same subscriber database for existing 3G clients and WLAN clients, allowing centralized billing and maintenance for different technologies. However, the new link AAA-HLR requires standardization. Loose coupling utilizes the common subscriber database without any user plane Iu interfaces, i.e., avoiding the SGSN and GGSN nodes. It allows for independent deployment and traffic engineering of WLAN and 3G networks. The common factor for all the proposed architectures in this category is the use of Mobile IP as the basic instrument for inter-system mobility and the high-level perspective of the integration process. In these solutions, interconnected networks are considered as independent networks concerning the handling of data traffic.

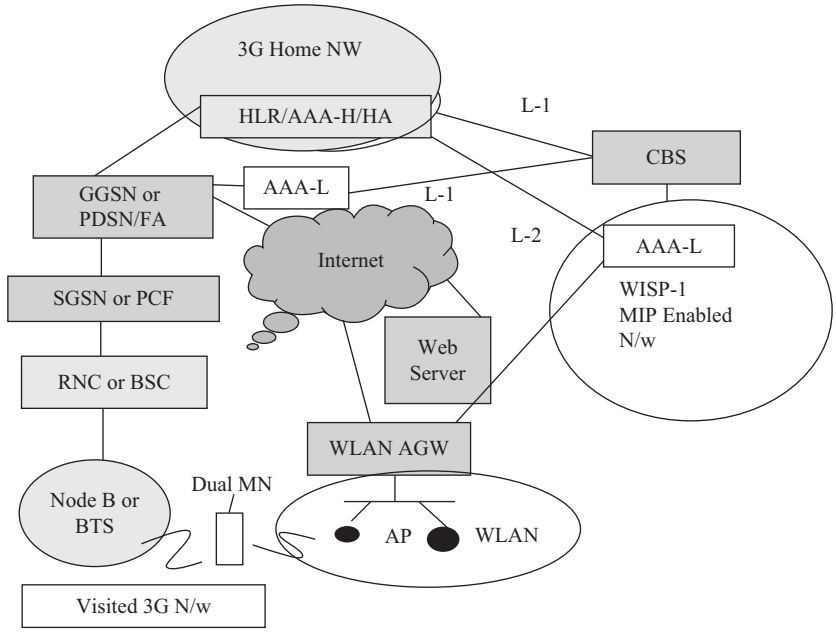


Fig. 15.5 *Loose coupling integration architecture*

In this architecture Foreign Agents (FAs) are placed at the borders of each access network to provide roaming capabilities. Moreover, HAs (Home Agents) are situated at the home networks along with authentication entities (AAA components for WLAN and HSS for UMTS authentication). If users have a different subscription to each network in case of SIM-based authentication for UMTS and login/password in WLAN then they cannot experience service continuity while roaming [2, IEEE survey]. On the other hand, a single subscription to one network with roaming privileges to another access network can help to avoid service disruption, as long as the different authentication entities are cooperating closely. Such cooperation has already been standardized by 3GPP [3].

There are two subclasses in LCIA.

(a) Open coupling at the CBS level: Here, **customer care and billing system** (CBS) is the common link between two independent WLAN and 3G networks (Fig. 15.5). This is purely an administrative integration.

(b) Loose coupling at common authentication and CBS: Common authentication requires a new interface between the AAA-HLR/HSS link and the Inter-Working Unit (IWU).

These inter-working functions are of two types:

(i) In 3G networks, the mobile application part is used to communicate with the HLR (Home Location Register), and UEs have USIM (UMTS SIM) cards for authentication.
- On the WLAN side, the UMTS-IWU (Interworking Unit) is required to support interoperability between WLAN's local AAA and the 3G core networks, HLR or HSS.

(ii) In 3G networks, DIAMETER/RADIUS is used to communicate with the home AAA-H in the 3G core-network (3G-CN).
- WLAN UE does not have USIM, so on the WLAN side, IETF-IWU is required to support interoperability between the WLAN's AAA-L and the AAA-H of 3G networks.
- A WLAN user requires specific NICs (Network Interface Card) in UEs if SIM-card based authentication is used by WLAN.

LCIA is better than TCIA in that it allows the Wireless ISPs (WISPs) to build up their independent service spots.

Figure 15.6 illustrates the comprehensive TCIA and LCIA together [4,5]. The point of integration and method of implementation depends on how to use AAA, QoS provision, protocol stack modification and support for RADIUAS/DIAMETER [6]. In the following section, a brief overview of AAA protocols for RADIUS and DIAMETER will be provided that will be needed in understanding the rest of the integration architecture and the role of AAA protocols.

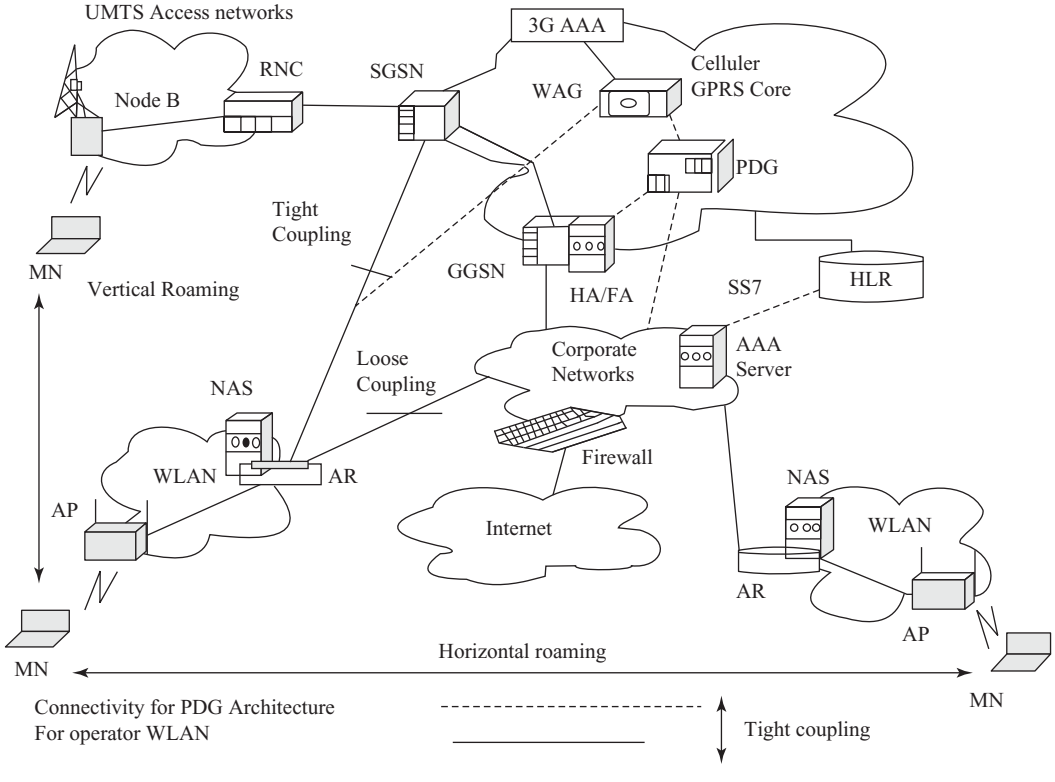


Fig. 15.6 *Comprehensive TCIA and LICA*

15.2.6 AAA Protocols: Radius/Diameter

The three As stand for authentication, authorization and accounting. This AAA is a very important building block used in construction of network security that helps to protect network operators and their customers from fraud, attacks, inappropriate resource management and loss of revenue. The word 'authentication' means something that is not false or fake, but based on trusty relationship. It mainly consists of two parts—one, providing proof of authenticity for the information that is being delivered and two, the other, act of verifying the proof of authenticity for the delivered information that is received. A client may be a device or a human being. When a client wishes to communicate to a network service provider, it needs to send the identity of the client along with a set of credentials. The network to verify that the identity actually belongs to the client then uses these credentials.

With the proliferation of public local-network providers like WLAN hot spots for passing users, many types of vulnerabilities will appear at various corners of the networks such as the following:

1. Customer–operator business and legal relationship may no longer exist.
2. Even the operator could trust the customer; it just needs to know the type of service or mapping of the user-device relationship. There may be multiple devices that belong to a team that the user belongs to, for example, government organisation, police department, etc.

Device-authentication credentials can be certificated or cryptographic keys loaded in the devices either by the manufacturer or by the network operators at the time of service negotiation. Message authentication ensures and verifies integrity of the received data. That is, the message comes from a legitimate source and is not altered by other parties. There is another term known as *mutual authentication***,** and it is in between the client and the server where two parties are simply peers. Each peer authenticates the other, either sequentially or in parallel.

Authorization is defined as the right to access a particular privilege, resource, communication link, information database (for example, accessing IEEE journals) or a particular computing machine. Another example of authorization is done in the case of the cellular prepaid card users. Every time the user sends request for a call, the network checks whether any credit is left or the service expiration date has been achieved.

The final 'A' is for accounting—the most important part of any customer-service provider relationship. It is not only the billing part but also includes auditing, cost allocation and trend analysis for future usage. Auditing verifies the correctness of an invoice submitted by a service provider. Cost allocation deals with the cost association to each of the telephony and data portions. Each of the parts may be processed at a different logic entity. In general, accounting is concerned with collection of information on resource consumption at all or specific parts of the network. The accounting data collected by the network device is carried by the accounting protocols over to the management entities responsible for each accounting application. The AAA protocols are considered application level protocols and hence must be carried by other communication protocols, including transport-layer protocols.

The mutual authentication between the client and the server is a special subcase of a more generic case of mutual authentication, where two parties are simply peers as opposed to client and server. In that case, each peer authenticates the other, either sequentially or in parallel. In a small network, the system administrator at the authentication terminal could configure an authentication server and they are collocated. For a large network, this is not practical. Many network points of presence acting as NAS (Network Access Server) are deployed and authentication must happen in a three-way party [6]. NAS controls communication into and out of the private network by intercepting messages between the supplicant and AAA server. NAS typically acts as a protocol dividing point. The AAA protocol must provide mutual authentication between the AAA server and the NAS. The most prominent AAA protocol is RADIUS (Remote Access Dial In User Service), designed to allow a NAS to forward a user's request and its credentials to the server. The response of the server will be carried back to the user through NAS. Thus,

RADIUS is a client–server mechanism, in which a NAS usually acts as a RADIUS client. The end user that is authenticated to the network through NAS is not the client in RADIUS. During the authentication procedure, the RADIUS client is responsible for passing user information in the form of requests to the RADIUS server and waits for the response from the server. Depending on the policy, the NAS may only need a successful authentication or further authorization directives from the server to open its traffic ports to the client's traffic.

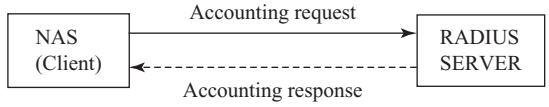


Fig. 15.7 *Basic RADIUS operation*

A NAS capable of supporting RADIUS accounting generates an accounting request at the start of the operation and sends it to the RADIUS server. This request message contains the type of service being delivered and the user information that the service is being delivered to along other things. Upon receiving a valid request, the RADIUS server adds an accounting record to its log and acknowledges the request by generating an accounting response. At the end of service delivery, the client will generate an accounting stop packet describing the type of service that was delivered. Both the accounting request and response messages are transported over UDP.

The **Extensible Authentication Protocol (EAP)** is designed to support a generic authentication protocol without requiring specific upgrades to the NAS. Since EAP relies on the backend server to understand and perform the actual authentication, RADIUS was extended to provide EAP support. When transmitting between the NAS and RADIUS servers, EAP messages are carried over the AAA protocol. The EAP-RADIUS framework allows the messages for EAP authentication schemes to be embedded inside RADIUS attributes and carried along RADIUS messages. This is illustrated in Fig. 15.8.

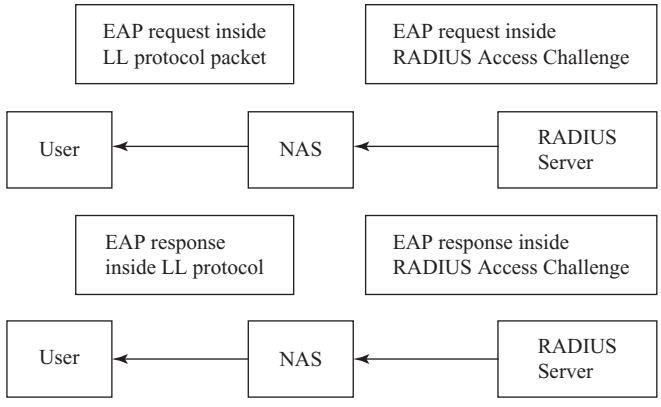


Fig. 15.8 *RADIUS interaction with EAP*

EAP is an extension of the PPP protocol. The IETF PPP extension group designed the EAP specifications in the form of RFC 2284 [EAP2284] as an extension of PPP. EAP was originally designed to support network access and authentication mechanisms in environments where IP messaging was not available. That is why it runs over data link layers, such as PPP and Ethernet-type links. EAP, by itself, does not perform the act of authentication but provides a mean for the negotiation between the user and the authentication server with NAS in the middle.

When it requires supporting for mobility and multi-domain operation, the RADIUS protocol is not very suitable. The RADIUS specifications provided by IETF do not provide any support for Mobile IP operation. Again, RADIUS suffers from security and reliability issues. The number of allowable attribute types in RADIUS is limited to less than 255, and the length of the attribute value fields are limited. This, along with the limited RADIUS message set, reduces its applicability as an AAA protocol in future mobile wireless networks. Thus, DIAMETER as a successor protocol to RADIUS, overcomes many of the RADIUS shortcomings. DIAMETER 3588 [7] defines most of the DIAMETER basic building elements such as the set of messages, attributes and the attribute structure. It also provides inter-realm operations in comparison to RADIUS.

The most important aspect of DIAMETER protocols is the concept of applications. A good example of a DIAMETER supporting application is Mobile IP. DIAMETER also specifies the security specifications from

end-to-end and hop-to-hop security mechanisms. AAA extension is needed for interaction with Mobile IP agents. The Mobile IP agents (HA and FA) use the AAA protocol to interact with the AAA server and are considered as AAA entities. The AAA protocol specifies the attributes and their semantics for carrying Mobile IP-related information between the AAA entities. Between the two most popular AAA protocols (RADIUS, DIAMETER), only DIAMETER provides detailed specifications to support Mobile IP. DIAMETER Mobile IP application defines new commands and command codes for interaction between mobile IPv4 agents and the AAA server. It also defines interaction for the variety of mobility and trust scenarios depending on the way mobile nodes register. Apart from this, DIAMETER defines the procedure to support Mobile IP handoff and specific attribute value pairs (AVPs) to transport mobile IP related information between agents and the AAA server. It supports dynamic home address assignment and home agents for the mobile node whenever needed, especially in a roaming scenario. It also provides guidelines to use NAI (Network Address Identifier) for the mobile node and routing of the DIAMETER messages towards the home AAA server (HAAA).

Interaction between DIAMETER and Mobile IP agents depend on the way the Mobile IP foreign agents are deployed. There are two types:

1. **FA–based care-of address (CoA):** When a foreign network deploys FAs and the mobile node uses FA based CoA during registration through the FA. This registration request is then forwarded to the HA. If the necessary Mobility Security Association (MSA) to support mobile IP authentication does not exist, the FA forwards the request to the AAA server (Fig. 15.9(a)).

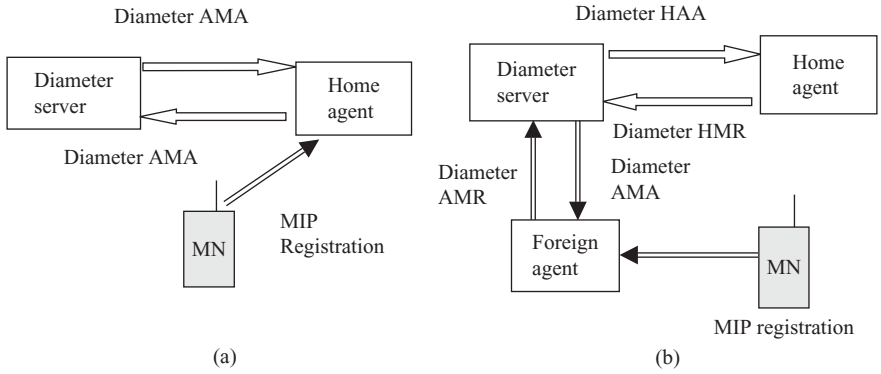


Fig. 15.9 *Diameter server and Mobile IP agent interaction (a) Without FA (b) With FA*

2. **Co-located CoA:** FAs are not deployed on the co-located CoA. So the mobile node directly sends a registration request to the HA. In turn, the HA forwards the request to the AAA server for the authentication of the mobile node (Fig. 15.9(b)). Within a foreign domain, the FA must contact its local AAA (LAAA) server, which in turn communicates with the home AAA (HAAA). Fig. 15.9 shows the DIAMETER server interaction with the mobile agents for the two cases discussed above.

   1. AMR (AA Mobile Node Request) message sent by the mobile agent to the DIAMETER server (HAAA if the mobile is in the home domain)
   2. AMA (AA Mobile node Answer) sent by the DIAMETER server to the mobile agent
   3. HMR (Home Agent MIP Request) message sent by the DIAMETER server to the HA to process the Mobile IP registration request
   4. HMA (Home Agent MIP Answer) sent back from HA to the DIAMETER server in response to HMR

SIM-based Authentication It is a very popular authentication mechanism and originates from the cellular GSM world, based on the Subscriber Identity Module (SIM). SIM-based authentication is basically a challenge/response mechanism with a 128-bit random challenge, called RAND, by the network authentication server. The cryptographic processor present on the SIM card generates a response (SRES) based on the

operator's specific algorithm and key Ki that is stored in the SIM memory. The authentication server also keeps a copy of the key Ki and performs the same operations on the RAND. The result is compared with SRES received from the SIM card of the user. If the result matches then the user is considered as valid and authenticate. Apart from the SRES, the SIM card also produces an encryption key Kc that is used for the encryption on the wireless link. In the GSM system, the groups (SRES, Ki, Kc) are together called triplets.

15.3 DESIGNING DUAL MODE TERMINAL

In WLAN and 3G interworking architecture, the 3G terminal should be able to access the WLAN. So, 3G terminals should have WLAN accessing capability. The radio technologies of 3G terminals and that of WLAN are different with respect to key issues such as link control, MAC protocol, coding, framing, transmitter and receiver. This leads to the issue of designing dual-mode 3G terminal equipment. In order to get access for high-speed data transfer, the cellular terminal equipment needs to have WLAN access capability as the primary goal. In a mobile scenario, if the terminal equipment accesses two different access networks (AN), it has to adapt the protocol stacks of the RLC/MAC layer of both APs. This can be solved in two ways. Firstly, the terminal equipment should have two different air interface cards, one for GPRS/UMTS (Cellular) and another for WLAN air interface. Secondly, one interface card can be configured on requirement basis to adapt to a particular access technology through a software-defined radio (SDR) system. But SDR technology has yet to come a long way to be used as a commercial product.

Dual-mode mobile terminal equipment (for 3G and WLAN) can be architecturally divided into two categories. One model is the Common LLC Layer Terminal Equipment (CLL-TE) where a common **GPRS LLC layer** works on lower layers of both GPRS and WLAN. Another model is the Common IP Layer (CIPL) terminal equipment, where a **common IP layer** works on the lower layers of both GPRS and WLAN. The RLC in GPRS controls a radio link access. The LLC sublayer maintains a logical link between MN and SGSN and a Data Link Control Identifier (DLCI), which comprises of a Service Access Point Identifier (SAPI) and a Temporary Logical Link Identifier (TLLI) for each link. The TLLI issues are controlled by the GMM sub-layer [3GPP TS 04.60 v8.23.0 (2004-05)].

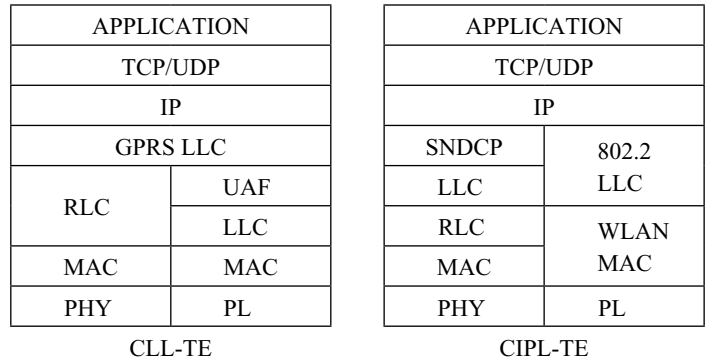


Fig. 15.10 *Protocol stacks for CLL-TE and CPIL-TE*

15.3.1 CLL-Terminal Equipment

In GPRS-WLAN networks, the LLC PDU can be transported by the WLAN radio system by inserting a User Adaptation Function between GPRS LLC and WLAN LLC layers [8]. A peer **GPRS Inter working Function (GIF)** in WLAN has to be implemented and it works like another access network for SGSN as shown in Fig. 15.11. The following functionalities have been outlined for roaming support.

Activation of the WLAN interface and routing area update message has to be sent to the GMM (GPRS Mobility Management) after association with WLAN, transfer of LLC PDU, discovering of MAC address of GIF and routing area identifier of WLAN.

Again, the GIF needs support paging procedure on the Gb interface. For downlink LLC PDUs received on the WLAN correlation between TLLI and 802 MAC, an address is required. Correlation between IMSI (International Mobile Subscriber Identity) and 802 MAC address for paging is also required.

This mode of terminal equipment is best suited for tight coupling because all cellular signaling between mobile nodes and networks are transported through WLAN. All advantages of tight coupling such as

security, handoff, sessions and service continuity are achieved. But this will have a severe impact on existing 3G cellular networks, as high-rate WLAN traffic will be injected in cellular core networks. In the tight-coupling method, one severe problem is that even the destination address is in the same WLAN, the packets travel through SGSN and GGSN. Moreover, this incorporates the modified implementation of **User Adaptation Function (UAF)**, terminal equipment and GPRS Interworking Function (GIF) at access networks. The mobile node uses all cellular control signaling with an Access Router (AR) in WLAN. Although it avoids the use of a SIM card reader and EAP protocol at the mobile node, yet it is less suited for loose coupling, as it needs a complicated implementation of inter-working functions at the AR. So, in loose coupling, this will incorporate a tremendous control signal overhead.

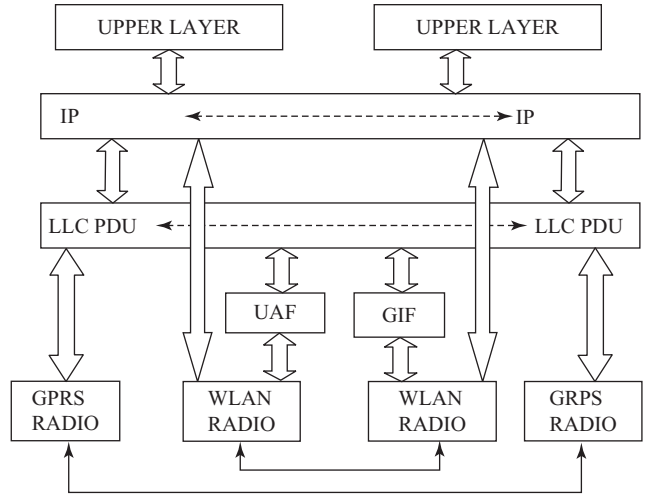


Fig. 15.11 *Transportation of LLC PDU through CLL-terminal equipment*

15.3.2 CIPL-Terminal Equipments

For this terminal mode, the IP layer is common to both the GPRS and WLAN terminal protocol stack as depicted in Fig. 15.12. For this architecture, the system complexity is transferred to the network side. As TCP/IP is the most promising and widely applied successful technology, the IP layer connectivity of WLAN and GPRS will support many other state-of-the-art technologies including Mobile IP. WLAN can be connected to corporate IP networks or the Internet, which is connected to GPRS through the Gi interface. In order to implement SIM-based authentication, the terminal equipment must be fitted with a SIM-card reader or smart card reader and control signaling can be transported by RADIUS/DIAMETER protocol over IP networks as given in Fig. 15.6. A Network Access Server (NAS) at the access router and an (RADIUS) AAA server in the core network are used. The RADIUS server uses SS7 network and protocol to communicate with HLR at the cellular network. The user is recognized as a host of WLAN and all conventional technologies are applicable without any modification in hardware as well as in protocol stack. This provides full support for Mobile IP and implementation of all IP heterogeneous networks.

The use of a CIPL terminal equipment in the tight-coupling method has been proposed in [1] where EAP is avoided and Local Mobility Management (LMM) functionality has been implemented (Fig. 15.12) in the application layer. Having recognized the SSID beacon from an Access Point (AP) in WLAN, the mobile node performs the power-up registration through GPRS. After that it associates with WLAN and the LMM (Link Layer Mobility Management) sends the routing area (RA) update message to SGSN through WLAN. The authentication shall take into account the user's subscription profile, and optional information about the WLAN Access Network (AN) operator's name and WLAN location, i.e., country, telephone, area code, city, etc. This information would enable the use of location-based authentication and authorization; location-based billing to customer care and location-based service offerings. At the terminal equipment, a complicated state-transition management policy is to be implemented for activation of the radio system of both GPRS and WLAN. The implementation of interworking functionality at AR through a Gb interface and one billing system is too difficult. The packet meant for destinations within the same WLAN will suffer excessive overhead, as it travels through GGSN.

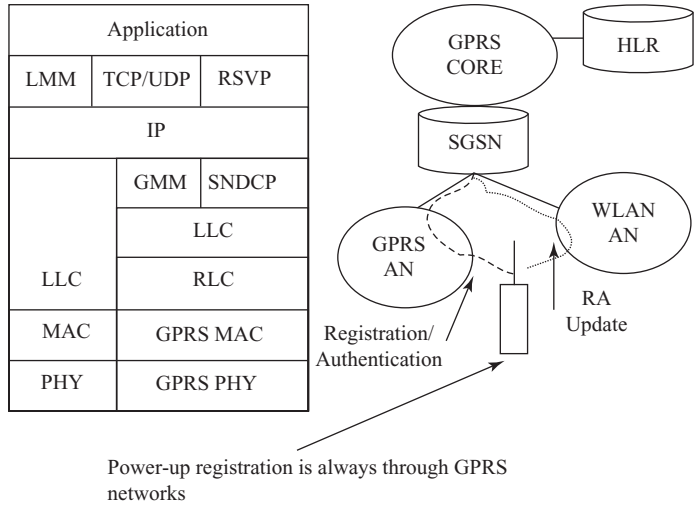


Fig. 15.12 *Protocol stack of CIPL terminal equipment used in TCIA*

15.4 IP-BASED LOOSE COUPLING

The protocol stack of 3G-terminal equipment and that of WLAN converging at the IP layer is best suited for loose coupling. 3G-WLAN heterogeneous network architecture [9-12], AAA handling and vertical hand-off mechanisms using IP-based loose coupling are gaining more importance for building a next-generation all-IP network. In the loose-coupling method, for centralized authentication, authorization and billing, the EAP-AKA/SIM (the GSM authentication and key agreement/SIM) is run at the terminal equipment, and NAS uses the RADIUS protocol for control signaling with the AAA server. The AAA server runs MAP (Mobile Application Part) with HLR over SS7 networks. The RADIUS server with a GPRS backend would provide a transparent authentication interface for roaming scenarios in a WLAN network. The GSM Authentication and Key Agreement (AKA) define an EAP type that allows a GSM/UMTS mobile client to be authenticated in WLAN networks. The EAP-SIM uses the SIM for authentication purposes against the GPRS networks. It does so by using the GPRS Security and Mobility Management (GMM) Protocol in order to perform GPRS attach/detach and authentication operation. Several loose-coupling approaches are discussed in the following subsections.

15.4.1 Gateway Approach

A gateway approach for the integration of GPRS and WLAN has been implemented using the common IP terminal equipment in the loose-coupling method in [13]. The primary issue in the integration of GPRS and WLAN lies in the integration of Mobile IP (MIP) with mobility management defined in GPRS. A gateway is a logical entity, which could be implemented standalone, or as an addition to the gateway GGSN, which connects the external networks as illustrated in Fig. 15.13.

A user might have his/her home networks in either GPRS or WLAN networks, but the gateway should be able to function like a Home Agent (HA) or Foreign Agent (FA). When dynamic addressing from the home PLMN (Public Land Mobile Network) or a visited PLMN is used, it is the responsibility of the GGSN to allocate and release the dynamic PDP address [14]. When an external PDN (Packet Data Network) address allocation is used, the PLMN may obtain a PDP address from the PDN and provide it to the MS (mobile subscriber) during PDP context activation, or the MS may directly negotiate a PDP address with the PDN after the PDP context activation procedure is executed. During the PDP context activation, if the PLMN is to allocate an external PDN address, the GGSN and PDN use DHCP and RADIUS for dynamic PDP address allocation and release. If the mobile node itself negotiates a PDP

address with the PDN after PDP context activation in case of external PDN address allocation, it is the responsibility of the mobile node and the PDN to allocate and release the PDP address by means of protocols such as DHCP or MIP.

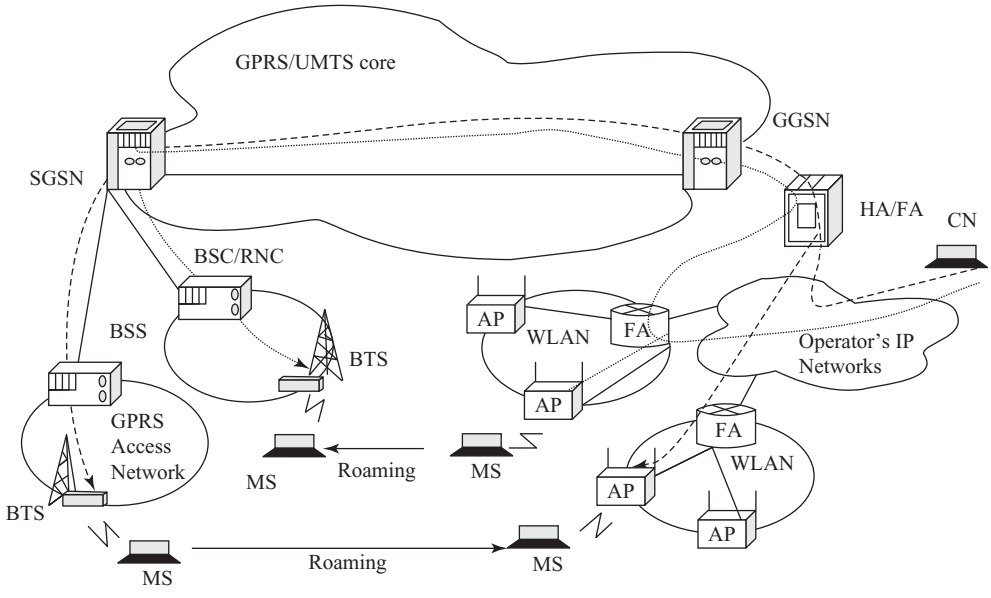


Fig. 15.13 *LCIA: Gateway architecture*

In case of DHCP, the GGSN provides the function of a DHCP relay agent and in case of MIP, the GGSN provides the function of an FA. For dynamic IPv6 address allocation, the GGSN informs the mobile that it shall perform stateful auto-configuration [18] by means of the router advertisements. For this purpose, the GGSN shall automatically and periodically send router advertisement messages towards the mobile node after a PDP context of type IPv6 is activated. Fig. 15.14 describes the MIP registration process in the gateway approach.

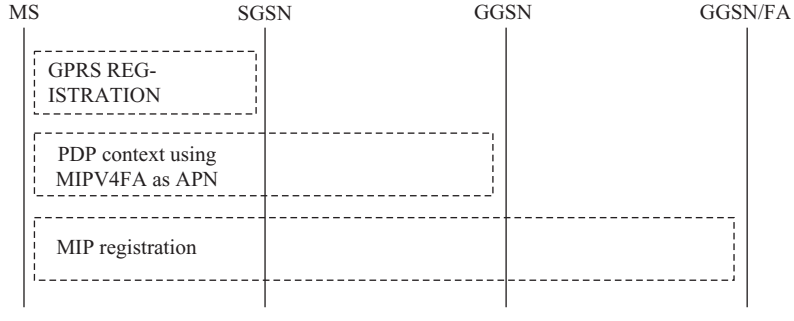


Fig. 15.14 *MIP registration through GPRS in gateway architecture*

15.4.2 Operator WLAN System

Another loosely coupled GPRS-WLAN network architecture is the Operator WLAN (OWLAN) system that combines the GSM subscriber management and billing mechanism with WLAN access technology and has been implemented using dual-mode common IP terminal equipment [9]. The OWLAN solution is available

for any wireless LAN terminal device that provides a GSM SIM-card reader and a defined operator WLAN signaling module. A single subscriber identity should be used in all access networks to enable smooth roaming and seamless service availability. The main design challenge was to transport the standard GSM subscriber authentication signaling from the terminal to the cellular site using the IP protocol framework. User data packets are directly routed to the IP backbone, which is used for accessing the public and private services.

WLAN can complement mobile operators' wide area GPRS and GSM service portfolios to offer cost-effective wireless broadband data solutions in indoor environments. The objective is to utilize mobile operator strengths of large GSM customer bases, cellular infrastructure investments, and roaming agreements in operator WLAN networks. The new development of the WLAN system architecture with the GSM subscriber management and billing system is called the OWLAN solution. The OWLAN network architecture is shown in Fig. 15.15.

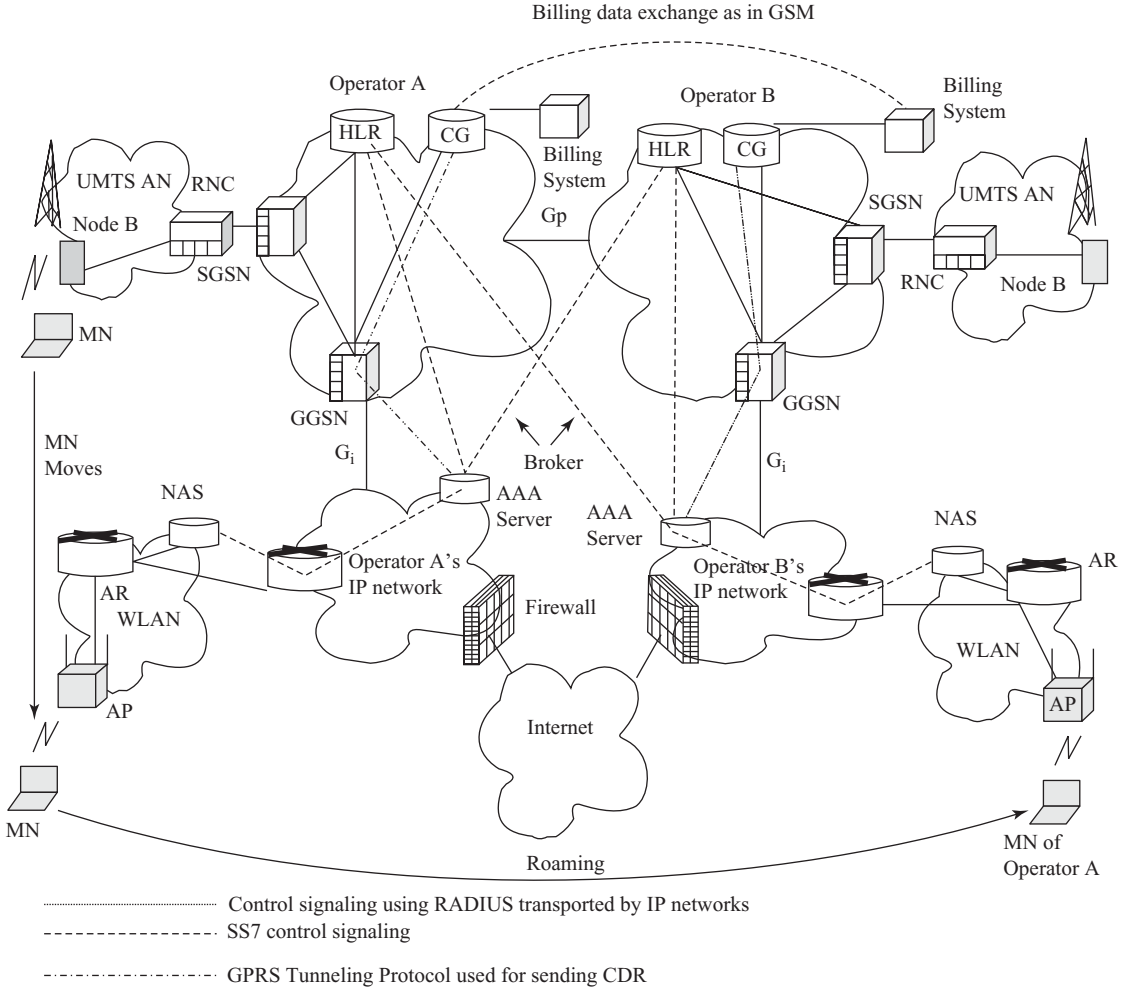


Fig. 15.15 *Operator's WLAN network architecture for GPRS/WLAN integration*

The OWLAN enables user roaming between different operator-access networks. The operator WLAN solution is available for any WLAN terminal device that has a GSM SIM-card reader and a defined

OWLAN signaling module. The OWLAN system maintains compatibility with the existing GSM/GPRS core-network roaming and billing functions, which minimises the number of modifications in the GSM core equipments and the standardization. OWLAN contains four main physical entities—AS (Authentication Server), AP (Access Point), AC (Access Controller) and MN (Mobile Node). In the OWLAN system, the control signaling data is transported to the cellular core. AC is the component responsible for routing data packets directly to the IP network backbone. OWLAN decreases the load of cellular networks.

The AAA server communicates with the access controller using the RADIUS authentication protocol. It receives the accounting data from the AC and converts it into the GPRS billing format. The authentication server sends the standard GSM authentication and signaling networks that connect the various operators. In the mobile terminal, a SIM-card reader, SIM-authentication software module and roaming-control module are necessary. SIM-specific signaling messages are transported using the IP protocol. Network Access Authentication and Accounting Protocol (NAAP) encapsulates GSM authentication messages inside the IP packets using UDP. This is illustrated in Fig. 15.16.

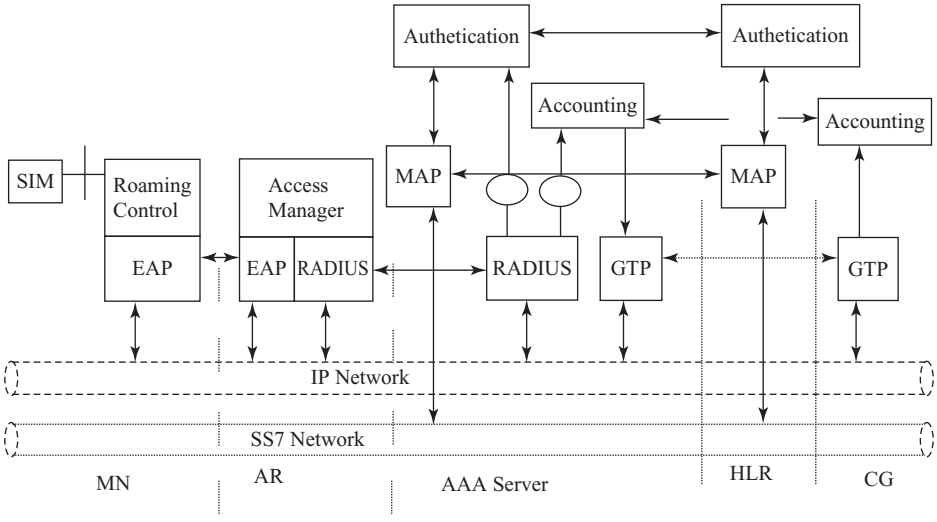


Fig. 15.16 *SIM-based control signaling for operator WLAN architecture*

15.4.3 Internet Roaming Architecture

The operator-oriented integration solution only covers public WLANs in hot spots, and not private WLANs such as office or residential WLANs. A WLAN-cellular integration solution called Internet roaming for corporate user roaming between these environments has been implemented in [10]. An Internet roaming system consists of a Virtual Single Account (VSA) server deployed on a corporate intranet, a Secure Mobility Gateway (SMG) deployed between the public Internet and the corporate intranet, and the **Internet Roaming Clients (IRC)** installed on the user terminals as shown in Fig. 15.17. Corporate users create a secure connection using a single sign-on authentication interface regardless of which wireless network, the computer is connected. Implementation of this scheme only involves the deployment of an SMG and a VSA server at proper locations on the intranet and installation of the IRC on the mobile computer of every user who needs secure mobile networking functions. The IRC is a suitably configured software program or a hardware interface card for mobile laptops or PDAs that users can install. The VSA server stores access credentials in a VSA record, which contains a user's single sign-on

VSA username and password, an intranet profile, a cellular profile, and several WLAN profiles. The SMG is deployed between the public Internet and the corporate intranet. It authenticates a user's computer with the help of the VSA server. The IP packets transmitted between the SMG and the mobile at its current location are encrypted and encapsulated. In context of Mobile IP, the intranet is the home network, the SMG is the home agent, and the IP address used by the user's computer is that computer's home address. By using mobile-IPsec packets, a single IP-in-UDP tunnel between an IRC and the SMG suffices for both security and mobility.

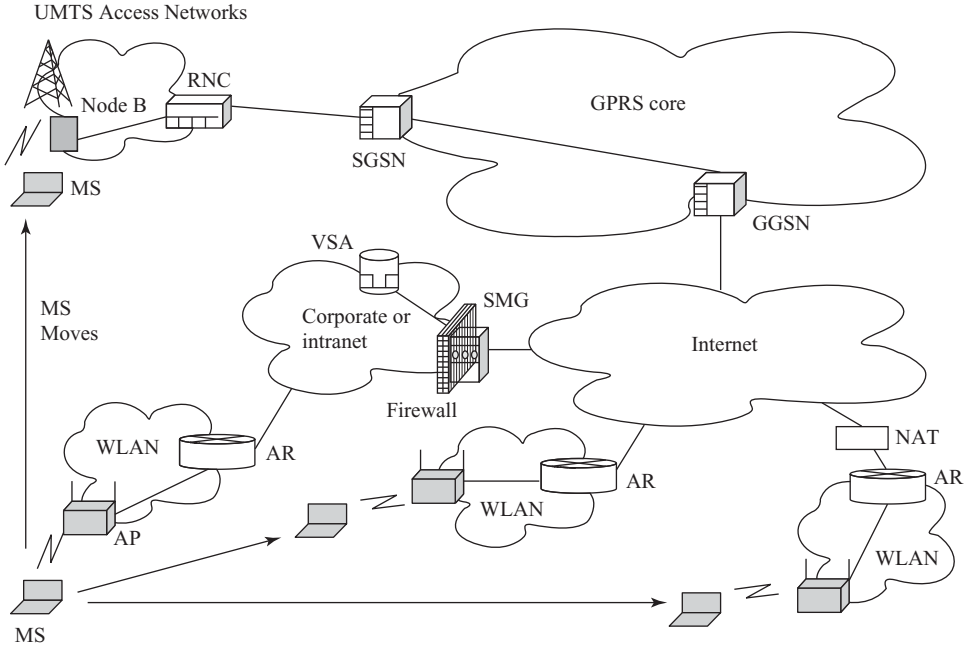


Fig. 15.17 *Internet roaming loosely coupled architecture*

15.4.4 A Global Architecture

A mobile device with both cellular phone and WLAN capability is gaining more importance in providing wireless Internet access services in hot spots such as airports, cafes, and bookstores. When the services of wireless LAN and cellular networks are integrated, the mobile station can move across those networks.

A global architecture (Fig. 15.18) for high-performance secure cellular/WLAN integration has been proposed in [12]. The cellular networks and base stations are 3G based and access points, which form a hot spot, are the attachment points that allow mobile stations to wirelessly access the network. A mobile station, as a member of its home network, is allowed to access the resources in the home network, and whenever it is within or outside the home network, a special server called the CA server is deployed in the integrated network. The CA checks authenticity for fixed nodes and networks upon request. The mobile station does not contact the CA server directly because of the large population size. The CA shares an independent secret key with the servers to which the mobile node is connected. The WLAN roaming is based on an AAA broker with a RADIUS proxy server. RADIUS is popular and easier to use for integrating hot-spot services into AAA-based cellular networks.

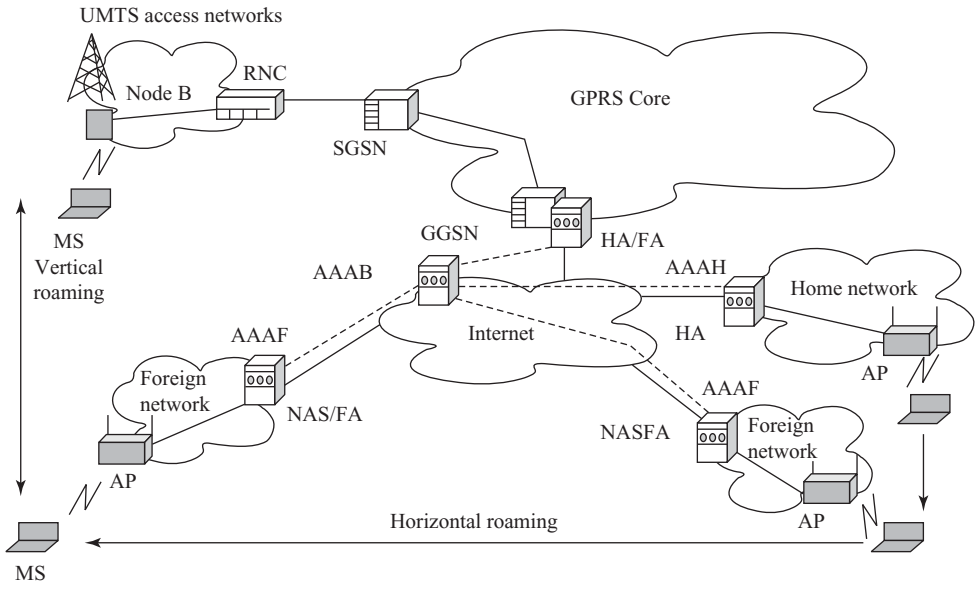


Fig. 15.18 *Global architecture*

The broker model is suitable for large-scale and commercial implementation, because a RADIUS server can simply have one simple security association or a present shared secret with the RADIUS proxy. The RADIUS server retrieves the remote server's domain from the user's request that includes the network access identifier (NAI), the user's name and the domain to which he or she belongs. Then it forwards the request to the remote server identified by the domain. The remote server also replies through the forwarding server. In order to provide IP mobility in the hot spot, the functionality of an FA is imposed in the NAS. The FA periodically sends advertisements with challenge packets and all mobile stations register via the FA. The challenge is a piece of data used to verify if the user device has knowledge of the secret key without sending it explicitly via a communication link. Figure 15.19 shows the authentication process for the global architecture.

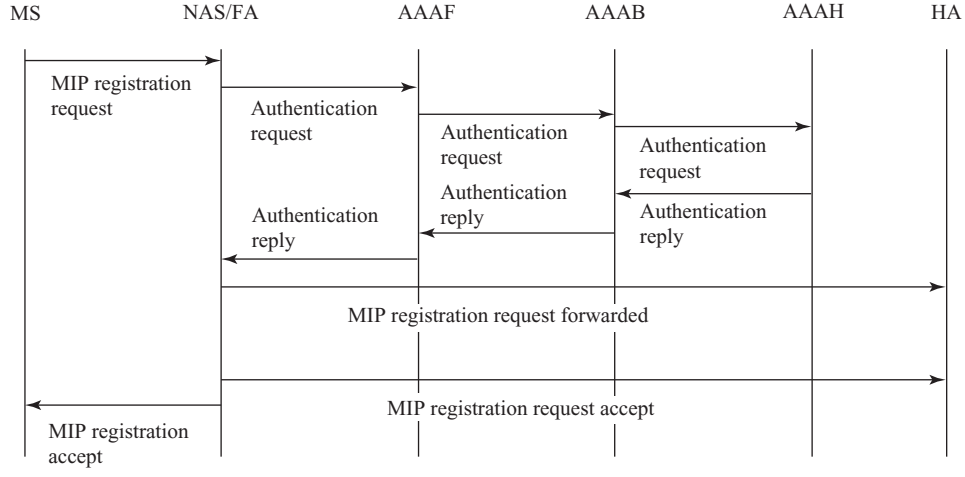


Fig. 15.19 *MIP-based control signaling for global architecture*

15.5 IP-BASED TIGHT-COUPLING ARCHITECTURE

The IP-based tight-coupling architecture [4,5] is shown in Fig. 15.20. WLAN is directly connected to a WLAN Access Gateway (WAG) in 3G cellular-core, and a separate Packet Data Gateway (PDG) is

connected to the external PDN through the new interface W_i. A key element in the architecture is the **3G AAA server**, which is a new functional component incorporated in a 3G PLMN (Public Land Mobile Network) in order to support interworking with WLANs.

The 3G AAA server in the 3G home PLMN terminates all AAA signaling with the WLAN and interfaces with other 3G components, such as the Home Subscriber Server (HSS), Home Location Register (HLR), Charging Gateway/Charging Collection Function (CGW/CCF), and Online Charging System (OCS). Both the HLR and HSS are basically subscription databases used by the 3G AAA server for acquiring subscription information for a particular WLAN subscriber. Typically, if an HSS is available, the HLR need not be used. The CGW/CCF and the OCS are 3G functional elements used to provide offline and online charging services respectively.

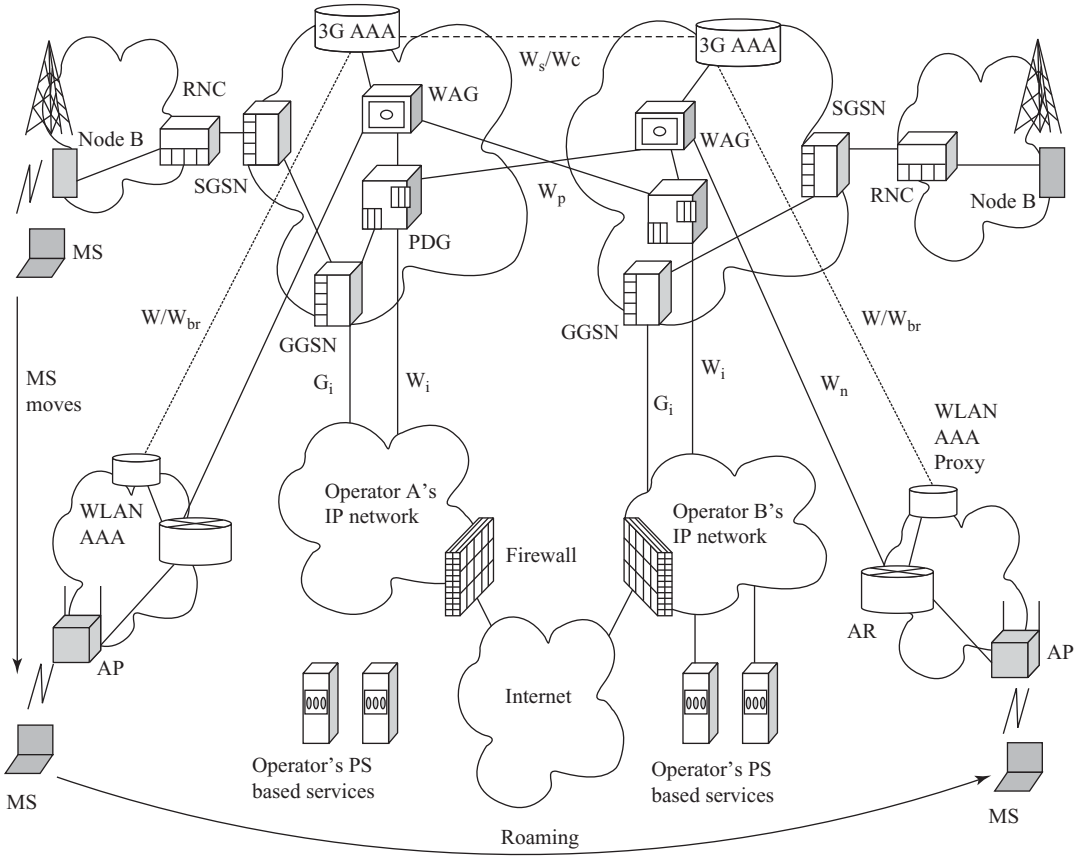


Fig. 15.20 *Packet data gateway architecture: TCIA*

The 3G AAA server can also route AAA signaling to/from another 3G PLMN, in which case it serves as a proxy and is referred to as a **3G AAA proxy**. In Fig. 15.21, the 3G AAA server in the 3G visited PLMN takes the role of a proxy that routes AAA signaling to/from the 3G AAA server in the MS's home PLMN. The 3G AAA proxy identifies the IP address of the 3G AAA server by using the network address identifier (NAI) sent by the MN.

There is also a WLAN AAA proxy that resides in the WLAN and routes the AAA messages to the appropriate 3G AAA proxy/server based on the NAI sent by the MN. As noted before, the user data traffic of WLAN MNs is routed through the WLAN to an external intranet/Internet. Although this traffic is routed by the WLAN, the 3G PLMN (both home and visited) can apply a specific policy to this traffic. For example, during the authorization phase the 3G AAA server may define a set of restrictions to police user data traffic,

such as 'do not allow FTP traffic' or 'allow HTTP traffic only to/from a specific IP address'. The user data traffic is routed from the MN's 3G home PLMN to a new component called a *packet data gateway* (PDG). User data traffic is also routed through a data gateway, referred as a Wireless Access Gateway (WAG), which in the case of roaming is located in the preferred 3G-visited PLMN. For AAA signaling, the MN conveys a W-APN (WLAN Access Point Network) to the 3G AAA server in order to explicitly indicate that access to a 3G PS-based service is requested. The W-APN is an identifier formatted as an Internet domain name that designates the PS-based service requested by the MN and possibly the operator through which this service is requested. In effect, the W-APN identifies the PDN that provides the requested PS-based service.

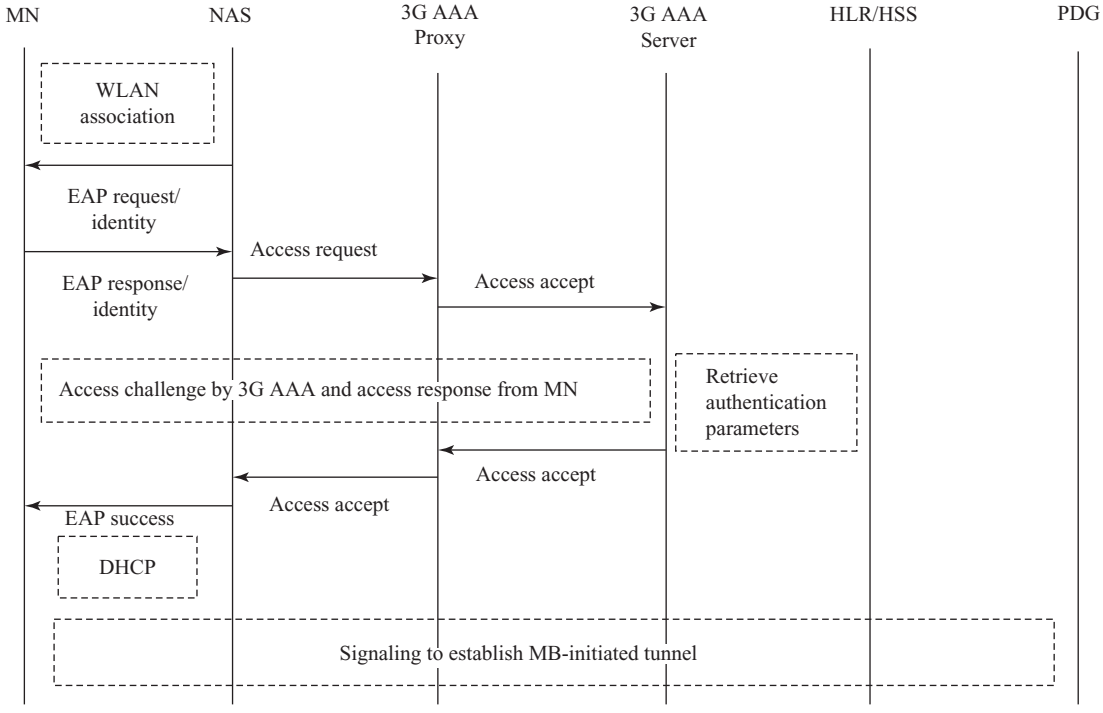


Fig. 15.21 *Control signaling in PDG architecture for MN initiated tunneling*

For user data transfer, either of two tunnelings can be established. One is **MN-initiated tunneling** that corresponds to the case when tunnel establishment is initiated by the MN itself, and hence originates at the MN and terminates at the PDG. Fig. 15.21 illustrates a typical signaling flow for this case. The only difference is that the W-APN (WLAN Access Point Network) is also carried from the MN to the 3G AAA server in order to explicitly indicate that access to a 3G PS-based service is requested. This W-APN is used by the 3G AAA server for authorization purposes and initiating tunnel establishment procedures if necessary. The second is **MN-Transparent Tunneling** that corresponds to the case when the tunnel is established without any MN intervention. An example of such a configuration is depicted in Fig. 15.22. In this case, the MN is roaming and uses PS-based services in the home PLMN. Two MN transparent tunnels are established—one from a WLAN access gateway (WLAN AG) to the WAG and another from the WAG to the PDG in the home PLMN. In this case, all MN outbound traffic is routed to the WLAN AG with normal IP routing, then it enters the first tunnel to the WAG, and finally reaches the PDG over the second tunnel. Note that the WLAN AG, WAG, and PDG need to make forwarding decisions based on previously installed information, which is communicated during the AAA phase. The PDG would also perform address translation if the MN's IP address were not assigned from the address space of the 3G-home PLMN.

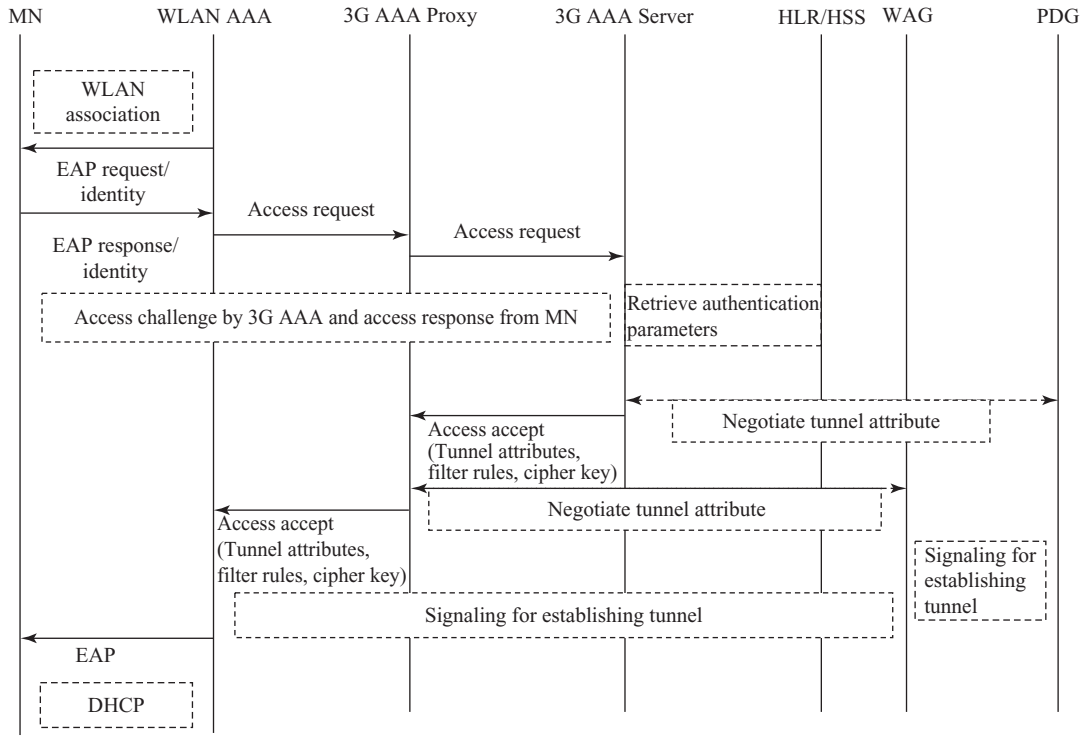


Fig. 15.22 *Control signaling for MN-transparent tunneling*

Based on the above discussions and summarizing all points. Figure 15.23 shows the cellular (GPRS/UMTS)/WLAN integration scenarios.

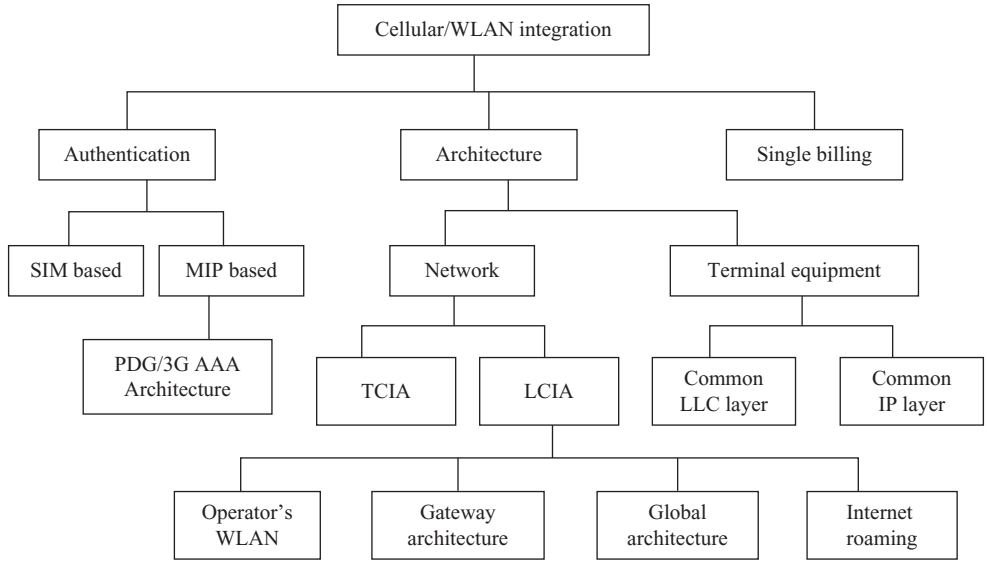


Fig. 15.23 *GPRS-WLAN integrated networks architecture classification*

15.6 HANDOFF IN INTEGRATED NETWORK ARCHITECTURE

Handoff is a phenomenon through which seamless roaming of mobile users is possible. For any integrated architecture (UMTS /WLAN), the mobile node roams from one access point (AP) to another within the WLAN area or from WLAN to UMTS and vice versa. This necessitates the handoff of two types—one is called vertical handoff, when the mobile node moves from WLAN to UMTS or UMTS to WLAN, and the other is called *horizontal handoff,* when the mobile roams within WLAN coverage from one AP to another or from one base station to another base station within the coverage of cellular (UMTS) networks.

The vertical handoffs from WLAN to UMTS are essentially different from those of UMTS to WLAN. Similar to the horizontal handoffs, the vertical handoffs from WLAN to UMTS are mainly initiated due to the out-of coverage area. As a WLAN user moves away from the WLAN access point, the received signal power decreases. When the MN detects that the received signal strength from WLAN is below a threshold, the MN will initiate a handoff request to UMTS. The UMTS may admit the hand-off request if it has sufficient resources to accommodate the request. Otherwise, the UMTS will drop the request. In this case, the MN is totally disconnected from the interworking system.

In contrast, the handoffs from UMTS to WLAN are triggered to seek low-cost or high-speed services or reduce the UMTS congestion. Since WLAN has rather small coverage and usually locates within a single UMTS cell, the MN requesting a vertical handoff from UMTS to WLAN is always within the coverage area of the UMTS. If WLAN accepts the hand-off request, the MN will break the connection with UMTS and start to communicate with WLAN. Even if WLAN denies the hand-off request, the MN can still retain the original connection with UMTS as it is still within the coverage of the UMTS cell. That is, the MN is always connected to the interworking system. Actually, there is no real blocking for the vertical handoffs from UMTS to WLAN.

Both loose and tight couplings use Mobile IP to manage handoff. However, the binding process in Mobile IP takes a long time and much signaling traffic. In case of ping-pong effect, where a mobile node requests frequent handoffs when moving frequently across the service area borders of the two systems, the Mobile IP cannot handle the handoffs timely due to the long processing time and overwhelming signaling traffic. In this section, two hand-off mechanisms—one for usual tight coupling and another for loose coupling—would be discussed. Handoff for the usual tight-coupling mechanism will be discussed in the light of Reference [1].

In GPRS, mobility context related to a mobile node's mobility within UMTS is stored in both the SGSN and the mobile node itself, and is called **GPRS mobility context** (GMM). Since the node roaming in WLAN remains connected to GPRS, both the SGSN and the node maintain the GMM. The node and access router maintain mobility context related to the node's mobility within the WLAN network, called **WLAN mobility context (WMM)**. Hence, the mobile node maintains two mobility contexts—GPRS and WLAN mobility contexts—while roaming in a WLAN network. The GPRS mobility context does not contain any mobility related information for WLAN roaming. The node connectivity in WLAN requires one additional state in the GPRS mobility context called WMM connected state (Fig. 15.24).

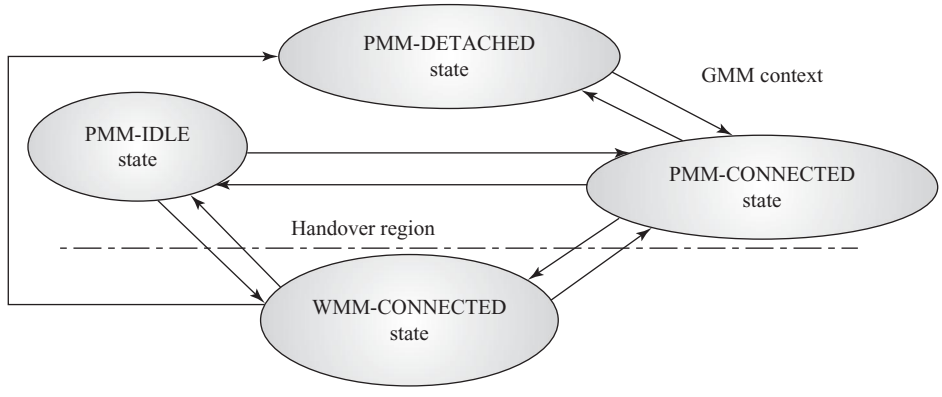


Fig. 15.24 *Mobility states for integrated networks*

All the three PMM states are GPRS context-related states. The WMM-connected state reflects the situation when the mobile is connected to the WLAN. The transition from the PMM-connected to WMM-connected state requires the inter-system handover procedure to run. In the GRPS mobility context, the MN remains in the WMM-connected state as long as it receives PS services through WLAN.

15.6.1 Inter-System Handover

Inter-system handover may be triggered due to the following reasons. The first two are operator specific while the later is user specific.

1. **Coverage expansion:** At the initial roll out of UMTS, coverage will be a limiting factor especially in the rural areas.
2. **Load balancing:** When the traffic in UMTS increases it would be advantageous to handover a certain part of the load to WLAN.
3. **QoS requirement:** The user may demand a higher QoS, for example, at a higher price. If it is available, the operator can then redirect it to WLAN capable of providing it.

The WLAN APs periodically transmit beacons, which are received by all powered-up MNs roaming within the AP coverage area, even those that are in power-saving mode. From the beacon signal, the MN detects its move from UMTS to WLAN and triggers inter-system handoff as described in Fig. 15.25.

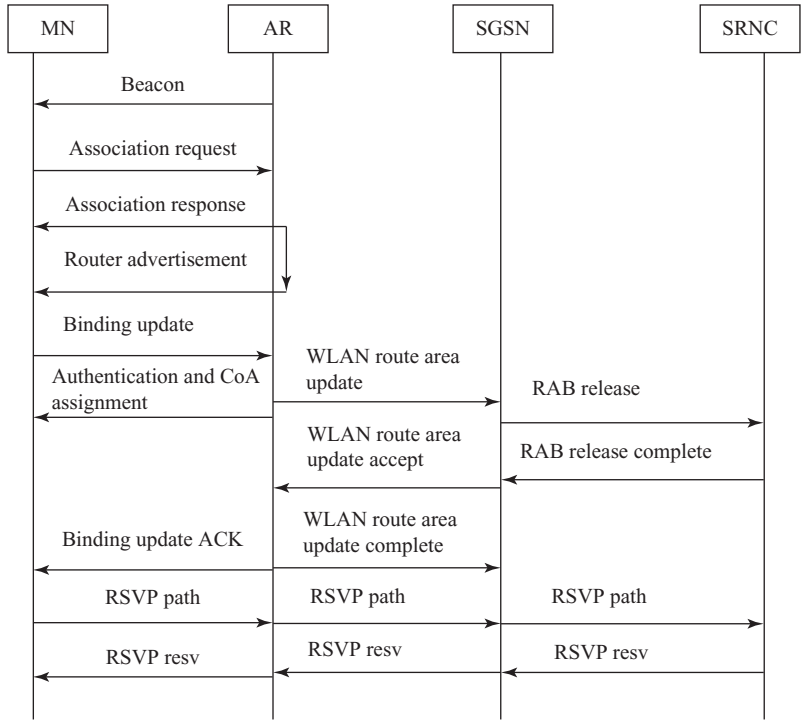


Fig. 15.25 *Handover mechanism from UMTS to WLAN*

Association-handshake is performed first by the MN to establish connection with WLAN AP through Layer-2 connectivity. The AP then generates a trigger to the AR after issuing association response to the MN. The AR sends a Router Advertisement (RA) to the MN to trigger the Layer-3 handover process.

The AR in this case need not wait for Router Solicitation (RS) from the MN before sending out an RA. In the next step, the MN performs Layer-3 handover. Binding of MN to AR is done by getting information from the RA about the IP address, etc., of the AR. The Binding Update (BU) contains the SGSN address and other UMTS-specific information to identify the mobile node and its mobility contexts in the SGSN. In response, the AR authenticates the MN with the AAA server and assigns a care-of address (CoA).

At this point, the AR starts the handover process with SGSN by sending a WLAN Route Area Update message. To break the radio resource access with the UMTS system, the SGSN first sends an RAB release message to the SRNC and then sends the acknowledgement to the Binding Update message. The SRNC releases the radio resource and channel for the packet data service and then sends the RAB release complete message to the SGSN. The SGSN is now ready to send the WLAN Route area Update Accept message to the AR and the GMM mobility context is changed to WMM mobility context. The WLAN Route Area Update message contains the SGSN PDP contexts list and the SRNC contexts. The SGSN first sends the WLAN Route Area Update Complete message, and then sends the Binding Update Acknowledgement message to the MN. This completes the UMTS-WLAN handover. Receiving the binding update ACK, the MN can now reserve the resources.

On the other hand, when the MN under WLAN moves away from its coverage, it detects missing beacons, the MN attempts to detect the AP by sending a probe. Failing to get a probe response, the MN detects that it moves out of the WLAN area and handover from WLAN to UMTS is triggered. The MN sends Route Area Update message to the SGSN through the UMTS interface. It indicates the disconnection to WLAN and contains the information for the last IP address of the AR association. The SGSN sends RSVP Reservation (RSVP Resv) Tear message to the AR, which causes it to a release radio resource to the MN. Figure 15.26 shows the WLAN-to-UMTS handover process.

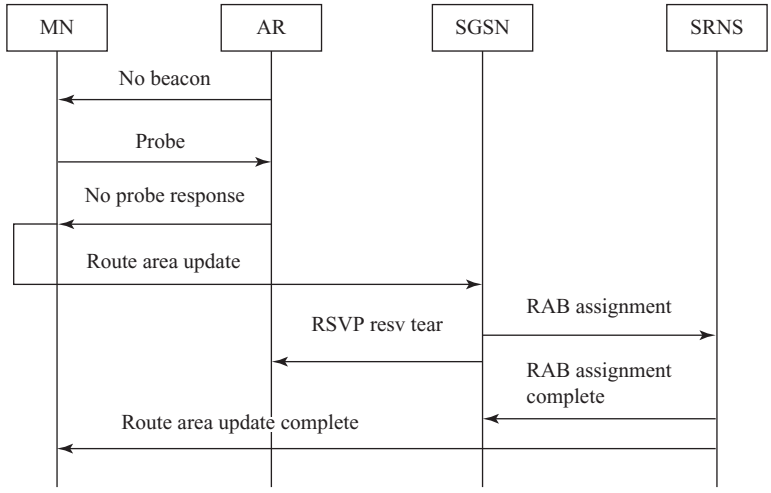


Fig. 15.26 *Handover process from WLAN to UMTS*

15.6.2 UMTS-WLAN Handover Process for Loose Coupling using Mobile IP

A loosely coupled architecture requires a new element, called WLAN gateway, that is connected to the Internet and not directly connected to the 3G-core network. WLAN gateway supports Mobile IP functionalities to handle mobility across networks as well as AAA services to interwork to 3G home networks and the AAA services. This enables 3G users to collect WLAN account and billing information.

An MN moving from a WLAN coverage area may experience degradation of signal and service and will have to perform handoff to 3G UMTS networks. If the WLAN MN finds no available AP after scanning, a hand-off notification is triggered. The hand-off procedures are initiated through the dual-mode interfaces of the MN.

The hand-off algorithm in the MN decides to disconnect from the WLAN and associate with the UMTS network. As the MN is Mobile-IP enabled, the FA is activated through the dual interfaces of the MN, and MIP is enabled. It gets a CoA for visiting the UMTS network as a foreign network assuming that to MN's home network is in WLAN. The Home Agent (HA) in the WLAN is informed about the new IP address through a Mobile-IP registration procedure and it performs a proxy ARP (Address Resolution Protocol) and intercepts the datagram. The HA encapsulates the datagrams and tunnels any packets arriving for the MN to the Foreign Agent (FA) of the UMTS network (here, GGSN). After decapsulation of the datagram, the MN

gets the packets through the UMTS network following the procedure for packet delivery.

The MN within the UMTS networks constantly monitors the air interface at periodic intervals to see the WLAN network availability. If it is found or is required due to high data-rate service, the handoff algorithm again initiates the association procedure to the newly discovered AP of the WLAN area.

After receiving the beacon signal from the AP, the MN discovers it WLAN network availability. The FA in the UMTS network is deactivated and updated by the Mobile IP and the home IP address is used now. Packets for MN are delivered through HA without going for a proxy ARP. Figure 15.27 illustrates the hand-off procedure from WLAN to UMTS network using Mobile IP.

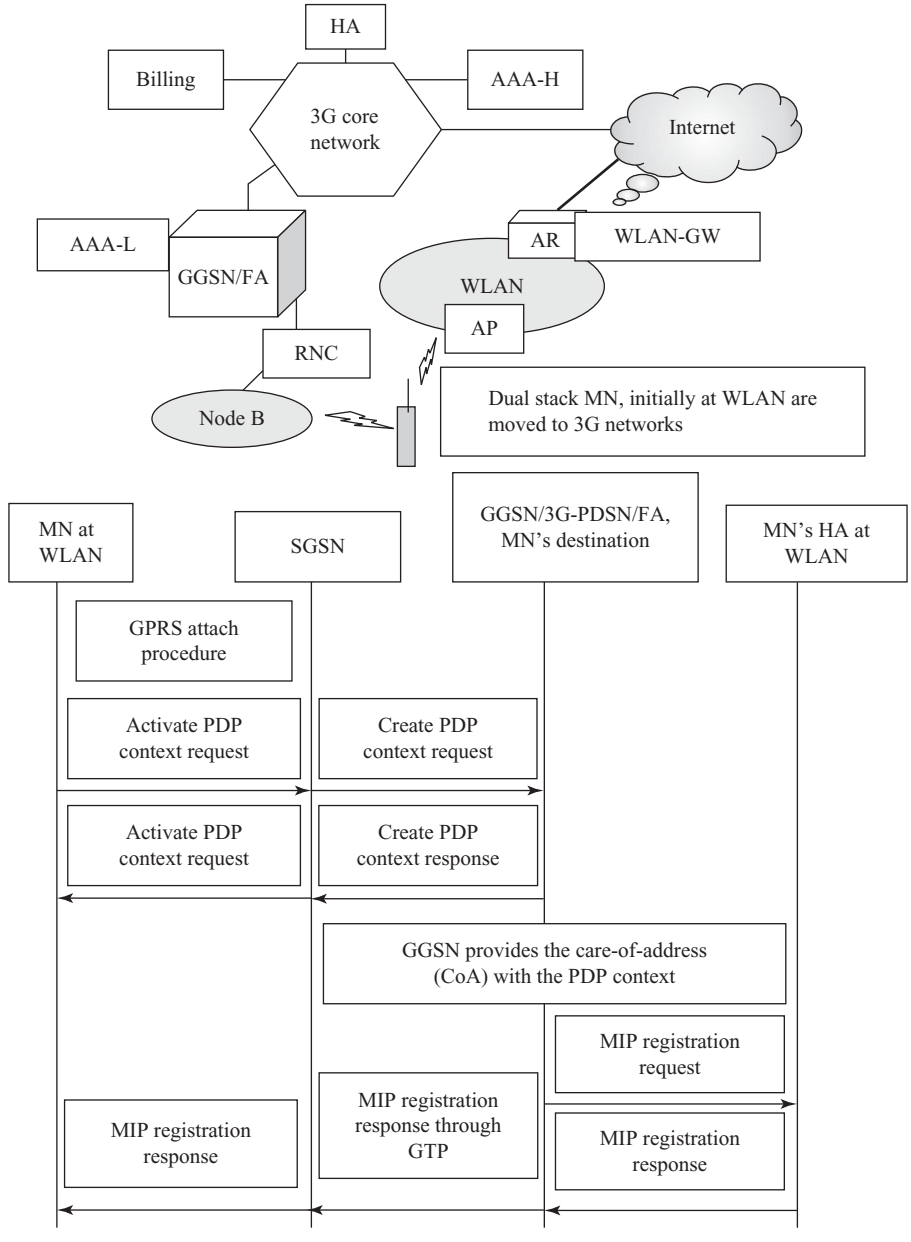


Fig. 15.27 *MIP-based handoff from WLAN to 3G (UMTS) network*

15.7 COMPARISON OVERVIEW OF DIFFERENT INTEGRATION ARCHITECTURES

1. **Billing issues:** The next-generation heterogeneous networks should fulfill a customer's choice of single billing from a single operator that provides all mobile services that the mobile subscriber owns. The billing methodology is not addressed in gateway and global architecture. A centralized billing system implementation is difficult when the user only subscribes either to GPRS or WLAN. Each home network maintains separate billing records. In the operator WLAN and PDG architecture, NAS at AR maintains the billing information and at the end of the session, this information is passed to the AAA server to tunnel the billing information to the charging gateway (CG) in cellular networks. In the roaming case, the billing data can be exchanged between two operator billing systems as in GSM. Implementation of billing systems is difficult in roaming while the MN is accessing a network other than the corporate networks.
2. **Mobility management:** Mobility management is a primary task for integration networks. In gateway architecture, the home network of a mobile node (MN) may be either in GPRS or in WLAN. When an MN subscribed to WLAN moves to GPRS, it performs a special MIPv4 FA registration with the GGSN and subsequently it performs MIP registration with the FA at GGSN. When an MN subscribes to GPRS, and roams to WLAN, it performs an MIP registration with the HA in GGSN, and subsequently the GGSN performs a PDP context deactivation for the same MN. The extra delay during handoff for a WLAN-subscribed MN moving in GPRS is introduced due to MIPv4 FA registration with GGSN before MIP registration with the FA at GGSN. This fact may lead to high hand-off latency. The extra delay during handoff for a GPRS-subscribed MN moving from GPRS to WLAN is introduced due to PDP context deactivation initiated by GGSN during registration with the HA at GGSN.

   In global architecture, each foreign IP network has its own AAA home/foreign (AAAH/AAAF) server based on the RADIUS protocol. A number of such servers are connected to the AAA broker (AAAB) server. In case of roaming the AAAB forwards the control packets to a remote home server identified by the domain. The MIP registration procedure is the same as in the gateway architecture.

   In operator WLAN and PDG architecture, the MN is subscribed to one operator, having services for both GPRS and WLAN. When the MN moves from GPRS to WLAN, a GSM-based authentication is performed. Authentication control signaling is transported using RADIUS/DIAMETWR protocol over IP networks. The AAA server does play the role of SGSN in regard to control signaling with HLR through circuit-switched SS7 networks. The implementation complexity of EAP over WLAN, NAS at AR, and AAA server at corporate/Intranet networks is more than that of HA/FA functionality at GGSN. Moreover, the SS7 network is costly.

   In the Internet roaming architecture, an MN subscribed to corporate networks is always reachable from other networks. The IRC is installed in terminal equipment and the SMG stores security association and location information in its memory for every wireless user. Its access capability in all networks is fully controlled by VSA and SMG installed in corporate networks. This architecture covers the private and residential WLAN at the cost of IRC at terminal equipment, VSA and SMG.
3. **Authentication:** In operator WLAN and PDG architecture, control signaling data is transported to the GPRS core for SIM-based authentication, and user data is directly routed to IP networks. After WLAN association with AP, the MN gets an IP address from the AR. The MN initiates the authentication procedure using NAAP or EAP protocol with AR, and the AR communicates for authentication with the AAA server using the RADIUS protocol in IP networks. The authentication server analyses the IMSI and verifies that the operators have a valid roaming agreement for WLAN services.

   Next, the authentication server sends an authentication query to the correct HLR using the GSM SS7 network. The corresponding HLR responds with the user profile and authentication triplets, and

the authentication procedure is completed in a normal way. In PDG architecture, the AAA server is implemented in cellular networks and it retrieves the authentication information from the HLR or HSS.

In gateway and global approach, the security function is performed during MIP registration. The home network of an MN may be in cellular networks or in WLAN. The Home Agent (HA) functionality for cellular networks is implemented in GGSN, whereas for WLAN, the HA in the access router is responsible for security functionalities.

In Internet roaming architecture, all cellular profiles and several WLAN profiles are stored in the VSN server, which works as a back-end server for the SMG. In all profiles, the authentication credential is encrypted via a key derived from the VSA password. A mobile-IPsec packet structure based on the existing IPsec and Mobile IP technology is used. The IRC authenticates the user to the network and SMG's interface and creates a mobile-IPsec tunnel to the SMG's interface.

Summary

In this chapter, the importance of cellular/WLAN integration architecture is discussed. The basic of two coupling modes, namely loose coupling and tight coupling architecture, and their different variants are discussed. Tightly coupled models require a technology-specific integration. Each of the networks being integrated must be modified to meet the requirements of integrating with another specific network, including protocols, interfaces and services. These modifications could be very complex and costly. Furthermore, tight coupling approaches are not much scalable due to their technology dependency, thus integration with future emerging technologies (e.g., IEEE 802.16, IEEE 802.20) could be even more complex and difficult. On the other hand, with a loosely coupled model, the architecture can be decoupled from specific standards and protocols, except IP protocols.

The CLL terminal equipment is best suited for tight coupling, and the CIPL terminal equipment is best suited for the loose coupling method. To avoid a large expensive SS7 network, the SS7 signaling traffic is transported through IP networks. For a CIPL terminal equipment, least modification of existing technologies; reuse of existing technologies such as EAP, RADIUS, SS7; and reuse of already established low-cost IP infrastructure are required, and it is best for all-IP-based next-generation heterogeneous networks, a step towards 4G networks.

References

[1] Jaseemuddin Md., *An Architecture for integrating UMTS and 802.11 WLAN Networks*, IEEE Symposium of Computer and Communication ISCC2003, Antalya, Turkey, Jun 30-July 3, 2003.

[2] Akyildiz, I.F., J. Xie, and S. Mohanty, *A Survey of Mobility Management in Next-Generation All-IP-Based Wireless Systems* IEEE Wireless Communications, pp. 16–28, Aug. 2004.

[3] *3GPP TS 23.234 V6.2.0, 3GPP system to Wireless Local Area Network (WLAN) Interworking*; System description (Release 6), Sept. 2004.

[4] 3GPP TR 23.934, *3GPP System to Wireless Local Area Netrwork (WLAN) Interworking; Functional and Architectural Definition (Release 7)*, Dec. 2007.

[5] Salkintzis, A.K., *Interworking techniques and Architectures for WLAN/3G integration toward 4G Mobile Data Networks*, IEEE Wireless Communication, pp. 50–61, June 2004.

[6] Nakhjiri, M., and M. Nakhjiri , *AAA and Network Security for Mobile Access, RADIUS, DIAMETER, EAP, PKI and IP Mobility*, John Wiley, 2005.

[7] Diameter 3588: P. Chlhoun, J. Loughney, E. Guttman, G. Zorn and J. Arkko, *Diameter Base Protocol*, IETF, RFC 3588, Sept. 2003.

[8] Salkintzis, A.K., C. Fors and R. Pazhyannur, *WLAN-GPRS Integration for Next-Generation Mobile Data Networks,* IEEE Wireless Communications, October 2002.

[9] Ala-Laurila, J., J. Mikkonen, and J. Rinnemaa, *Wireless LAN Access Network Architecture for Mobile Operators*, IEEE Communications Magazine, pp. 82–89, November 2001.

[10] Luo, H., Z. Jiang, B-jo Kim, N K Shankaranarayanan, and P. Henry, *Integrating Wireless LAN and Cellular Data for the Enterprise*, IEEE Computer Society, pp. 25–33, March-April 2003.

[11] Faccin, S.M., P. Lalwaney, B. Patil, *IP Multimedia Services: Analysis of Mobile IP and SIP Interactions in 3G Networks*, IEEE Communications Magazine, January. 2004.
[12] Shi, M., X. Shen, and J.W.W. Mark, *IEEE802.11 Roaming and Authentication in Wireless LAN/Cellular Mobile Networks*, IEEE Wireless Communications, pp. 66–75, Aug. 2004.
[13] Chen, J., H. Lin, *A Gateway Approach to Mobility Integration of GPRS and Wireless LANs*, IEEE Wireless Communication, Vol 12, April 2005.
[14] Bettstetter, C., H. Vogel, and J. Eberspacher, *GSM Phase 2 + General Packet Radio Service GPRS: Architecture, Protocols, and Air Interface*, IEEE Communications Surveys, Third Quarter, Vol. 2 No. 3, 1999.
[15] Politis, C., K.A. Chew, N. Akhtar, M. Georgiades, and T. Dagiuklas, *Hybrid Multilayer Mobility Management with AAA Context Transfer Capabilities for All-IP Networks, IEEE Wireless Communications,* pp. 76–88, Aug. 2004.

Questions for Self-Test

15.1 In tight coupling, the WLAN network acts as another 3G-access network where WLAN traffic is directly injected into the 3G cores.
a. True b. False

15.2 IP-based loose coupling integration method requires least modification in either terminal equipment or in networks.
a. False b. True

15.3 SGSN may be a suitable integration point in tight coupling rather than GGSN, as mobility context is maintained by SGSN.
a. True b. False

15.4 Loose coupling allows for independent deployment and traffic engineering of WLAN and 3G networks.
a. True b. False

15.5 The overall traffic of a GPRS core network increases for
a. TCIA b. LCIA

15.6 Complementary benefits are taken in any integration architecture.
a. False b. True

15.7 Mobility management is the key issue in the 4G heterogeneous architecture.
a. False b. True

15.8 EAP is the extension of PPP protocol.
a. True b. False

15.9 The two main standardization bodies for integration are _____ and _____.

15.10 Operators would seek the _____ data rate of _____ and _____ coverage for 3G in the integration architecture.

15.11 Seamless mobility among heterogeneous networks need to solve the _____ hand-off problem

15.12 The physical and data link layers are kept intact in integration architecture while incorporates _____ and _____ stacks.

15.13 AAA stands for _____, _____ and _____.

15.14 _____ is defined as the right to access a particular privilege, resource and information database.

15.15 Discuss the five basic implementation issues for cellular-WLAN integration.

15.16 Compare and contrast the 3G and WLAN services with respect to integration.

15.17 Integration at GGSN for TCIA may increase hand-off delay. Explain.

15.18 Explain the operation of very tight coupling integration and point out the pros and cons.

15.19 Highlight the benefit of integration with respect to operator's service.

15.20 What are the challenges in designing dual mode terminals? How many types of terminal varieties are there?
15.21 Which mode of terminal is best suited for tight coupling and why?
15.22 What are the key features of SIM-based authentication?
15.23 Point out some benefits of cellular-WLAN integration architecture.
15.24 What is the most suitable point of integration in the GPRS/UMTS –WLAN integration architecture?
15.25 Discuss the role of the AAA server in the integration architecture.
15.26 How do RADIUS and DIAMETER handle authenticity within a heterogeneous network environments?
15.27 What are the different types of loose-coupling integration architecture?
15.28 Mobile IP is the most suitable protocol for gateway LCIA. How?
15.29 Describe how security functionalities are maintained in global architecture by the AAA server.
15.30 Highlight the pros and cons of different loose coupling architectures in a tabular form.
15.31 Discuss the two different types of terminal models thought for integration networks.
15.32 How does intersystem handover occur in a tight coupling integration?
15.33 Explain the hand-off operation for loose coupling architecture using Mobile IP.
15.34 Describe the different steps of MIP-based handoff from UMTS to WLAN system.
15.35 Show the different classifications of cellular-WLAN integration described in this chapter.

Overview of WiMAX Technologies: Broadband Wireless Communication

16

Introduction

The word WiMAX stands for **W**orldwide **I**nteroperability for **M**icrowave **A**ccess. It is the new revolution in the area of wireless broadband services. The combination of Internet with the broadband wireless access of WiMAX will change the face of worldwide communication by broadcasting the Internet to every possible walk of life, whether a person is in office, home or on the move. The unbelievable expansion in both Internet and wireless access during the past several years has generated the demand for high-speed Internet access through wireless networks. So, simultaneous growth for broadband services is underway. The combining effect of broadband services in the wireless domain is the result of WiMAX–the Broadband Wireless Access (BWA). BWA enables application services for real-time audio and video streaming, video conferencing and also Voice-over IP (VoIP) communication. There are two types of broadband wireless accesses. One is the traditional fixed broadband wire line concept using wireless as the transmission technology, thought of as the alternative of DSL (Digital Subscriber Line) or the cable modem. This is known as the fixed WiMAX. The second type is the broadband wireless known as mobile broadband which supports roaming–the concept of Mobile WiMAX. IEEE 802.16d is the standard for fixed WiMAX and 802.16e is the mobile WiMAX standard.

In this chapter, we shall discuss the basic concept of broadband wireless access technology, its general architecture and the service types offered by WiMAX.

16.1 EVOLUTION OF BROADBAND WIRELESS

Broadband is the capacity to deliver Internet access with a continuous 'always on' connection and the ability to both receive and transmit digital content or services at high speeds [1]. With broadband, services such as data, voice, and video, commonly known as multimedia, can be delivered together as one packet. Broadband Wireless Access (BWA) technology based on the IEEE 802.16 family of standards delivers high data rate over long distances. It is an interesting alternative to wired solutions such as cable networks, DSL link in the last mile, and as a WiFi hot-spot backhaul, cellular backhaul and optical backbone extension in the middle mile.

Broadband transmission can be divided into three types—wired broadband, wireless fixed broadband, and mobile broadband. Wired media involves transmission via copper, coax and optical fiber. But wireless media involves transmission via radio and optical links. Fig. 16.1 shows the different broadband transmission technologies. The wireless broadband system uses one or more broadcasting antennas and receivers to provide broadband services. In case of a wireless fixed broadband network, broadcasting antennas and receivers are fixed.

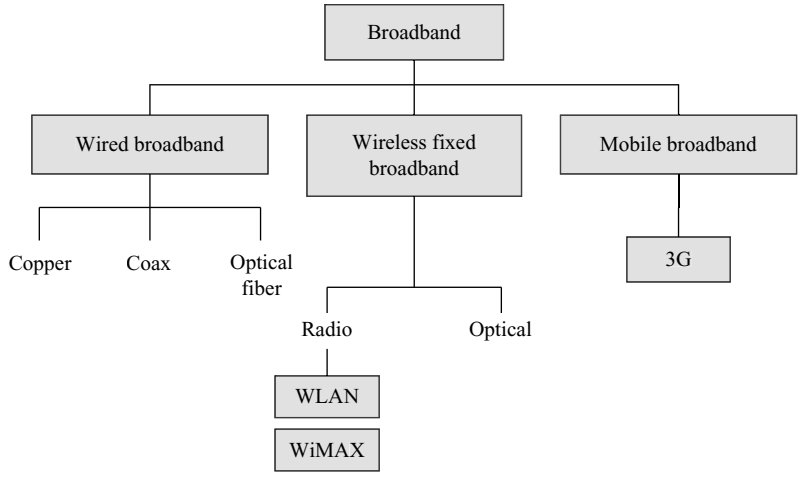


Fig. 16.1 *Different broadband technologies*

WiMAX is based on the IEEE 802.16 standard. WiMAX aims to provide business and consumer wireless broadband services on the scale of the Metropolitan Area Network (MAN). WiMAX is emerging as a broadband access technology with several advantages such as rapid deployment, high scalability, and low maintenance and upgrade costs [2]. The 802.16 will provide a cost-effective alternative to existing solutions based on very expensive leased-line services. Broadband wireless access technology has evolved through four stages as explained below.

1. **Narrowband WLL (Wireless Local Loop) Systems:** WLL is the first alternative for wireless access deployed for voice communication and is quite successful in the voice telephony market in India, China, Brazil, Indonesia, etc. This WLL system was based on DECT (Digital Enhanced Cordless Telephony) and CDMA (Code Division Multiple Access) technology. Some of the Wireless Internet Service Providers (WISP) offered Internet services in the unlicensed frequency band of 900 MHz and 2.5 GHz. Deployment was limited to small areas within a town.
2. **First-Generation LOS (Line-of-Sight) Broadband Systems:** This system began for higher frequencies at 2.5 GHz and 3.5 GHz bands to support high-speed access. In the late 1990s, MMDS (Multichannel Multipoint Distribution Services) at 2.5 GHz band was deployed for broadband wireless services. It was mainly used for broadcasting of video services in rural cable TV networks. But it could not compete with satellite TV. In September 1998, the Federal Communications Commission (FCC) allowed relaxed facilities for two-way communication in the United States. In the first-generation, fixed-type of broadband wireless services, long (more than 100 feet) high-power transmitters were deployed to provide LOS coverage to the wireless cable operators. Each cable operator needed to install his own antenna system at the outdoor environment in the direction of LOS with the base transmitter. The coverage area was approximately 30–35 miles under a single base station but the capacity was limited.
3. **Second-Generation NLOS (Non-line of sight) Broadband Systems:** The limitations of LOS were removed in the second generation. High-end digital signal processing is used for better system performance in the multipath environment and deployment architecture is almost cellular-like. IEEE 802.16 developed a standard in the year 1998 to standardise wireless metropolitan area networks (WMAN) at 10–66 GHz band for high-speed transmission. In December 2001, the IEEE 802.16 group developed a standard for WMAN with single carrier-modulation techniques for the physical (PHY) layer and a MAC (medium access control) using TDD/FDD (time division /frequency division duplex).

4. **Standard-based Broadband Wireless Systems:** The first IEEE 802.16 standard, approved in 2001, is in the 10–66 GHz range for line-of-sight wireless broadband services. In order to overcome the disadvantage of line-of-sight links, IEEE 802.16a, completed in 2003, is in the 2–11 GHz band for non-line-of-sight wireless broadband services, and OFDM (Orthogonal Frequency Division Multiplexing) is used as the part of physical layer. IEEE 802.16d [3], approved in 2004 and named IEEE 802.16-2004, is designed for fixed wireless communications. The new IEEE 802.16e standard [4] extends the 802.16d standard and provides mobility support in cellular deployments, and it was completed in December 2005. It specifies the scalable OFDM for the PHY layer and modifications to the MAC layer to accommodate high-speed mobility. The WiMAX forum has been formed to provide interoperability among the IEEE 802.16 standards.

16.2 SPECTRUM ALLOCATION

Spectrum allocation is a very important aspect for wide deployment of WiMAX for future broadband applications. The operating frequency band may influence the coverage and achievable data rates. In order to ensure that resulting 802.16-based devices are in fact interoperable, an industry consortium called the WiMAX Forum was created. The WiMAX Forum develops guidelines for frequency spectrum, the PHY layer to be used and some other parameters with the focus for interoperability between different vendor products. This WiMAX forum has identified some of the most likely frequency bands at 2.3 GHz, 2.5 GHz, 3.5 GHz and 5.7 GHz.

The **licensed 2.3 GHz band** is deployed in South Korea for WiBro (Wireless Broadband) services. This band is also available in Australia, New Zealand and United States. The FCC imposed a strong restriction for tight out-of-band emission in this band because of the Digital Audio Radio Services (DARS) adjacent to this band (2.320–2.345 GHz).

The bands between 2.5–2.7 GHz are the licensed bands allocated in the United States, Canada, Mexico, Brazil and some part of Southeast Asian countries. This is a very promising band for wireless service in the United States. But in many countries there is a restriction for fixed and two-way communications. The FCC allowed two-way communication for this band in the year 1998.

The **licensed 3.5 GHz band** is the primary band allocated for fixed wireless broadband services in many countries across the globe. International allocation for this band is 3.4–3.6 GHz. But there are other 3.3–3.4 GHz and 3.6–3.8 GHz bands also to be allocated for newer applications. The bandwidth availability varies from country to country. This band does not allow for mobile broadband services because of heavier radio propagation losses at this band.

The most interesting WiMAX application is the **unlicensed 5.25–5.85 GHz band**. This band may be the probable WiMAX deployment band for rural, low population density areas. The operator may coordinate frequencies for this higher bandwidth band, but it is very difficult to use in a mobile environment because of the high-frequency radio propagation loss. 5.725–5.850 GHz is the most attractive band for WiMAX application.

There is also the 80-MHz available unlicensed bandwidth at 2.4-GHz band for WiMAX application. But WiFi is already being used in this band, and WiMAX spectrum allocation in this frequency range thus will not be very suitable.

16.2.1 Frequency Bands at a Glance

10–66 GHz Licensed Bands (802.16)

1. Line-of-Sight (LOS) is required and multi-path is negligible.
2. Channel bandwidth is of 25 MHZ or 28 MHZ.
3. Raw data rate is 120 Mbps.
4. This single carrier modulation air interface is known as Wireless MAN-SC air interface.

5. Suits for point-to-multipoint (PMP) access serving applications from small office/home office (SOHO) through medium to large office applications.
6. In the 802.16 MAC, every BS dynamically distributes uplink and downlink bandwidth to the SS using Time Division Multiple Access (TDMA).
7. The MAC supports Time Division Duplex (TDD) and Frequency Division Duplex (FDD) scheduling for efficient operation.

2–11 GHz (802.16a) Licensed Band

1. LOS is not required.
2. Supports mesh architecture and Automatic Repeat Request (ARQ).
3. Multi path may be significant.
4. Uses orthogonal Frequency Division Multiplexing (OFDM).
5. Data rate can be up to 100 Mbps in each 20-MHz channel.

License Exempted Band (5–6 GHZ)

In addition to the 11-GHZ licensed band, the PHY and MAC introduce mechanisms such as Dynamic Frequency Selection (DFS) to detect and avoid interference.

16.3 WiMAX, WiFi, OPTICAL FIBER, AND 3G

WiMAX in combination with WiFi, optical fiber and 3G can provide broadband services to anyone, at any time, and from anywhere. Among various wired broadband technologies, broadband services using traditional telephone lines and cables are less expensive and they are useful in providing broadband to the last mile. But the installation of telephone lines and cables is very difficult in crowded geographical areas such as cities. Also, broadband services based on traditional telephone lines and cables are not available in rural areas or developing countries. WiMAX and WiFi may be an attractive alternative (WiMAX as WiFi hot-spot backhaul) in these areas. Local area network-oriented technologies such as WiFi are much more simple technologies, and are easily deployed. They focus on small coverage areas, usually inside buildings, where a very high bit rate can be achieved.

Optical fiber is lighter, covers greater distances and offers greater bandwidth than copper or coax. It is robust and has a long depreciation period. However, installing and maintaining optical fiber networks is more complex and costly. In addition, fiber optics is characterized by poor geographic scalability and very high initial investments. So, broadband services using optical fiber are uneconomical in the last mile. It is economical to use WiMAX as a backbone in the middle mile.

1. **WiMAX Versus WiFi:** Both WiFi and WiMAX are good at providing high data rate wireless packet services; however their coverage capacities are quite different. This enables the cooperation between WiFi and WiMAX—WLAN be the 'last hundred metres' access and WiMAX be the 'last hundred kilometres' access. The most important cooperation between WiFi and WiMAX is the deployment of WiMAX as a backhaul for WiFi AP's [5].
2. **WiMAX Versus Optical Fiber:** For remote areas WiMAX can be used as optical fiber extension [5].
3. **WiMAX Versus 3G:** WiMAX is known for its long-distance transmission and large capacity, so it can be complementary to fiber MAN in places where fiber cannot reach. 3G networks adopt a combined wireless access and fiber inter-connection scheme. For areas without fiber MAN or with no need for a temporary fiber base station, WiMAX is quite capable in providing the inter-connection among 3G base stations [5].

In respect of performance capacities, there are prominent differences between WiMAX, WiFi and 3G. 3G systems have a fixed channel bandwidth, but WiMAX defines a flexible deployment of a selectable channel bandwidth from 1.25 GHz to 20 MHz. So, throughput capacities of WiMAX are varied depending on the channel bandwidth.

Both WiFi and WiMAX use OFDM modulation scheme, whereas 3G is based on CDMA technology. The spreading for high data rate is very difficult in a CDMA system. WiMAX can achieve high spectral efficiency than CDMA-based 3G systems because of the use of multiple antenna (MIMO) systems. Because of the easy use of frequency diversity and multiuser diversity in OFDM systems in WiMAX, it offers higher peak data rates, greater flexibility, higher average throughput and system capacity.

3G is more expensive than the cost of WiMAX. But regarding the mobility aspect, 3G is more suitable as mobility is an integral part of the system; whereas in WiMAX the mobility feature is added on the fixed designed WiMAX systems. Table 16.1 shows the transmission capacity and range of WiFi/WiMAX/3G.

Table 16.1 *Transmission capacity and range of different wireless broadband technologies*

	Transmission capacity	Range
WiFi (802.11g)	Up to 54 Mbps	Up to 300 ft
WiMAX (802.16d)	Up to 75 Mbps	Up to 40 km
3G	Up to 2.4 Mbps	Typical 1–6 miles

16.4 IEEE 802.16 STANDARD ARCHITECTURE

The broadband wireless access systems are released as a standard IEEE802.16, which provides long range and large bandwidth. The 802.16 standard is created for broadband wireless access in order to offer high-speed capacity, low cost and scalable solution to extend the fiber-optics backbone. The 802.16 is defined as wireless MAN Air Interface for Fixed Broadband Wireless Access Systems. This provides network access to homes, small businesses and commercial buildings as an alternative to the traditional wired connection.

The IEEE 802.16 architecture [6] consists of two kinds of fixed (non-mobile) stations—Subscriber Stations (SS) and a Base Station (BS). The BS regulates all the communication in the network. The wireless MAN provides access to a subscriber station communicating with the central radio base station. The 802.16 supports two types of architecture—Point to Multipoint (PMP) architecture and mesh architecture. While the initial standard gives a maximum range of 8 km, the current standard of 802.16d in the family increases the range up to 40 km. The 802.16 can provide data rates of up to 75 Mbps in a 20-MHz bandwidth [5].

16.4.1 Point-to-Multipoint Architecture

In PMP architecture, a central base station handles multiple independent sectors simultaneously (Fig. 16.2). The BS controls all the communication in the network, i.e., there is no peer-to-peer communication directly between the subscriber stations. The BS is the only transmitter operating in this direction, so it transmits without the requirement of coordinating with other stations.

On the downlink, data to Subscriber Stations (SS) are multiplexed in Time Division Multiplexing (TDM) fashion. The downlink is generally broadcast. An SS can accept only that part of the downlink subframe, which is specified for that SS. The connection in the downlink direction is either unicast or multicast. The SS checks the Connection Identifier (CID) in the received Protocol Data Unit (PDU) and retains only the relevant PDU's address in them.

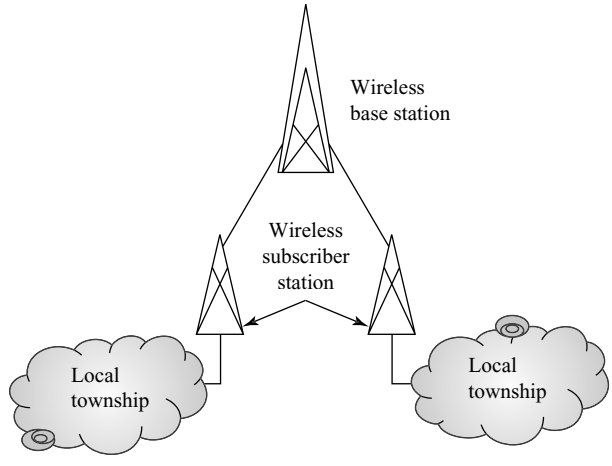


Fig. 16.2 *Point-to-multipoint architecture*

In the uplink direction, however, several SSs compete for access, so a Time Division Multiple Access (TDMA) technique is used where bandwidth and time slots are dynamically allocated based on demand. Uplink connections are always unicast. The SS may transmit only if the Base Station grants the permission after receiving a request from the user. This is very important. An SS requests uplink bandwidth on a per-connection basis. The BS grants bandwidth to a SS as an aggregate of grants in response to the per-connection request from SS.

In PMP architecture, each SS has a 48-bit MAC address, which is unique for equipment identification. A 16-bit CID identifies each connection. Upon entering the network, the subscriber station is assigned three management connections in each direction. These are

1. The basic connection that is used for the transfer of short time-critical MAC and radio link control (RLC) messages.
2. The primary management connection is used to transfer longer, more delay-tolerant messages such as those used for authentication and connection set-up.
3. The secondary management connection is used for the transfer of standard-based management messages such as Dynamic Host Configuration Protocol (DHCP), Trivial File Transfer Protocol (TFTP) and Simple Network Management Protocol (SNMP).

16.4.2 Mesh Architecture

The mesh mode of operation introduced in the IEEE 802.16a is based on the forwarding of data, received by one node to another. It allows creating a multipoint-to-multipoint network architecture. Nodes beyond the coverage area of the BS can be reached with this mode of operation, if the remote SS is in the coverage area of another SS.

In Fig. 16.3, *B*, *C*, *D* and *E* are SSs, which can directly communicate with the BS '*A*'. The subscriber stations *G*, *F* and *H* can establish the communication with the BS indirectly through the other SSs like *B*, *C*, *D* and *E*. These subscriber stations are called remote subscriber stations.

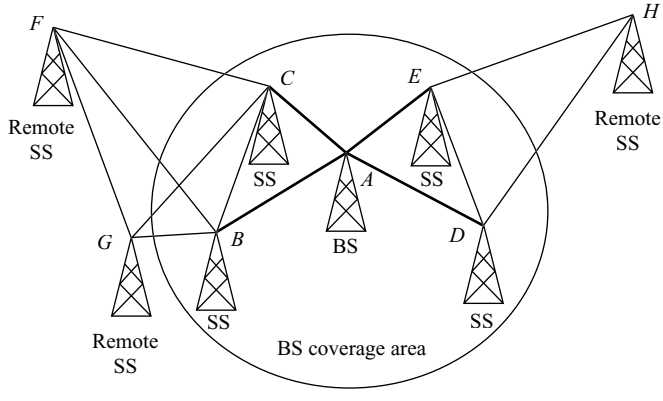


Fig. 16.3 *Mesh architecture*

In the PMP mode, traffic only occurs between the BSs and SSs, while in the mesh mode traffic can be routed through other SSs and can occur directly between SSs. This is the main difference between the PMP and mesh modes. Within a mesh network, a system that has a direct connection to backhaul services outside the mesh network is termed a mesh BS. All the other systems of a mesh network are termed mesh SS.

The three other important terms of mesh systems are neighbor, neighborhood and extended neighborhood. The stations with which a node has **direct links are called neighbors**. Neighbors of a node shall **form a neighborhood**. A node's neighbors are considered to be one-hop away from the node. An extended neighborhood, in addition **contains all the neighbors** of the neighborhood [3].

Each node shall have a 48-bit universal MAC address that uniquely defines the node from other equipments. For authorization to the network, each SS gives a request to the mesh BS, and then the mesh BS may issue a 16-bit node ID. This node ID is used for identifying the nodes in normal operation and is transferred in the mesh subheader in both unicast and broadcast message.

In the local network in the mesh mode, each node shall assign an 8-bit ID for each link it has established with the neighbors. This Link ID is transmitted as part of the CID.

16.5 OVERVIEW OF WiMAX PHY

In this section, we will provide a very brief overview of the IEEE 802.16 WiMAX PHY. This IEEE 802.16 standard defines both the medium access control (MAC) layer and physical (PHY) layer [3]. The **central BS** exchanges MAC protocol data with an individual SS. The link from BS to SS and SS to the mobile node would use a different physical link.

The WiMAX PHY layer is based on **OFDM** (Orthogonal Frequency Division Multiplexing). It is the transmission technique used to provide high-speed data, video, and multimedia communications generally used for various broadband communications (WiFi, DSL, WiBro etc). OFDM transmission is NLOS (Non-Line of Sight Propagation) that can handle multi-path radio propagation environment. OFDM converts a high-rate data stream into a number of low-rate streams that are transmitted over parallel, narrowband channels that can be easily equalised [12].

OFDM supports multi-carrier modulation schemes. The high-rate data stream is divided into several lower bit-rate data stream and each stream is on modulated separate carriers, called the subcarriers. The subcarriers are orthogonal to each other over the symbol duration. It avoids the need for non-overlapping subcarrier channels in order to eliminate inter-carrier interference. In order to completely eliminate inter-symbol interference (ISI), the OFDM uses guard interval between the symbols larger than the expected multipath delay spread as shown in Fig. 16.4. The spectra of different modulated carriers overlap, but each carrier is in the spectral nulls of all other carriers.

OFDM signal is equivalent to Inverse Discrete Fourier Transform (IDFT) of the data sequence block taken N number of subcarriers at a time. Thus, implementation of OFDM is easy by the use of Inverse Fourier Transform (IFFT) and Fast Fourier Transform (FFT). But the size of FFT in an OFDM should be designed very carefully to take care of the multipath delay spread and Doppler shift. For a given bandwidth, selecting a large FFT size would reduce subcarrier spacing and increase the symbol time, thus protecting against multipath delay spread. The multi-access scheme using OFDM where different subcarriers are partitioned among the multiple users are known as OFDMA. This technique is used in mobile WiMAX.

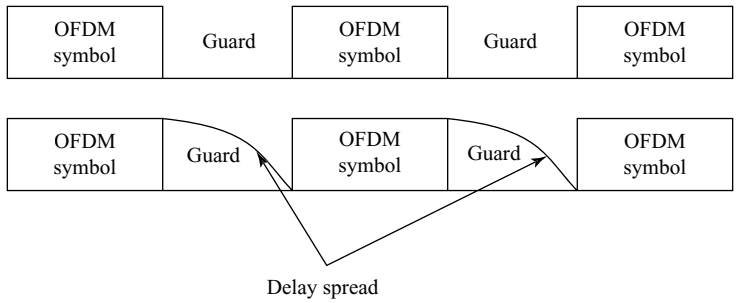


Fig. 16.4 *OFDM symbol and guard channel*

Grouping of N data symbols into a block is known as OFDM symbol which has a period $T = NT_s$, where T_s is the data symbol duration. In between two OFDM symbols, there is a guard channel to make each OFDM symbol independent after going through the wireless channel as illustrated in Fig. 16.4.

As long as the delay spread τ is less than the guard time T_g, the OFDM symbols will interfere among themselves at the receiver.

OFDM has high peak-to-average ratio (peak amplitude of the emitted signal/average amplitude) that creates non-linearity and clipping distortion in signals. This is because an OFDM signal is the superposition of N sinusoidal signals on different subcarriers. On average, the emitted power is linearly proportional to N. If the signals on the subcarriers add up constructively then the amplitude of the signal is proportional to N and thus the power is proportional to N^2. Moreover, OFDM signals are susceptible to phase noise and frequency dispersion that need to be taken care for implementation time [12].

For fixed WiMAX (IEEE 802.16d), the PHY layer is a 256-FFT based OFDM, whereas, mobile WiMAX (IEEE 802.16e) uses a flexible OFDMA-based physical layer where FFT size can vary from 128 bits to 2048 bits [6].

Design of the 2–11 GHz PHY is driven by the need for NLOS operation. The IEEE 802.l6a/d standard defines three different PHYs that can be used in conjunction with the MAC layer to provide a reliable end-to end link. These air-interface specifications are the following:

1. **Wireless-MAN-SCa:** A single-carrier modulated air interface.
2. **Wireless-MAN-OFDM:** A 256-carrier OFDM scheme. Multiple access of different subscriber stations (SSs) is time-division multiple access (TDMA)-based.
3. **Wireless-MAN-OFDMA:** A 2048-carrier OFDM scheme. Multiple access is provided by assigning a subset of the carriers to an individual receiver [6], so this version is often referred to as OFD Multiple Access (OFDMA). Of these three air interfaces, the two OFDM-based systems are more suitable for non-LOS operation due to the simplicity of the equalization process for multi-carrier signals. Of the two OFDM-based air interfaces, 256-carrier Wireless-MAN-OFDM is in a favorable situation to the vendor community for reasons such as lower peak-to-average ratio, faster FFT (Fast Fourier Transform) calculation, and less stringent requirements for frequency synchronisation [6].

For **fixed WiMAX OFDM PHY** among 256 FFT, 192 subcarriers are used for carrying data, 8 are used as pilot subcarriers for channel estimation and synchronization and the remaining parts are used as guard bands. The subcarrier spacing varies with channel bandwidth as the FFT size is fixed, thus affecting delay spread. For maximum delay spreads about 25% guard time can be used—16 μs at 3.5 MHz operating channel and 8μs at 7 MHz channel.

In **Mobile WiMAX OFDMA-PHY** FFT size is variable from 128 to 2048. The FFT size is increased when available bandwidth increases. But the subcarrier spacing always remains fixed to 10.94 kHz to keep a balance between the delay spread and Doppler spread requirement for fixed and mobile environments. This maintains the OFDM symbol duration fixed which is the basic resource unit. This subcarrier spacing can support a delay spread of up to 20 μs and a mobility of 125 kmph at 3.5 GHz operating frequency. When channel bandwidth is 1.25 MHz, 5 MHz, 10 MHz and 20 MHz respectively, the subcarrier spacing of 10.94 kHz corresponds to 128, 512, 1024 and 2048 FFT.

16.5.1 Subchannelizations

As mentioned earlier the BS controls all the communication in the network, i.e., there is no peer-to-peer communication directly between the SSs. The communication path between SS and BS has two directions—uplink channel (from SS to BS) and downlink channel (from BS to SS). The downlink channel, defined as a direction of data flow from the BS to the SSs, is a broadcast channel, while the uplink channel is shared by SSs. Time in the uplink channel is usually slotted (mini-slots) called by time division multiple access (TDMA), whereas on the downlink channel, the BS uses a continuous time-division multiplexing (TDM) scheme.

The entire subcarriers are divided into several groups of subcarriers known as *subchannels*. The fixed WiMAX standard OFDM-PHY allows 16-limited subchannelization in the uplink direction. 1,2,4,8, or all sets can be assigned to a subscriber station (SS) in the uplink. Uplink subchannelization in fixed WiMAX allows SS to transmit for **a fraction (1/16) of the bandwidth allocated** to it by the base station (BS). Mobile WiMAX subchannelization is based on OFDMA-PHY, and allows subchannelization both for uplink and downlink. Subchannels are formed by the minimum frequency resource unit allocated by the BS. Different subchannels may be allocated to different users as a multiple access scheme, hence the name OFDMA.

The frequency spectrum for subcarriers may be contiguous, or pseudo-randomly distributed for providing more frequency diversity that is useful for mobile WiMAX. The PUSC (Partial Usage of Sub-carriers) is mandatory for all mobile WiMAX implementations. The subchannelization scheme based on contiguous subcarriers in WiMAX is called band Adaptive Modulation and Coding (AMC). It allows the system designer to exploit multiuser diversity depending on the frequency response. The overall system capacity increases in multiuser diversity. Adaptive modulation and coding increases the overall system capacity as it allows trade offs-between the throughput and robustness of the channel.

16.5.2 Adaptive Modulation and Coding

The IEEE 802.16a/d standard defines seven combinations of modulation and coding rates that are used to obtain the various data rates and robustness depending upon the channel and interference conditions. The base station scheduler can take into account the channel quality of each user's uplink and downlink and assign a modulation and coding scheme for maximizing throughput with the available SNR (signal-to-noise ratio). Table 16.2 summarizes the different modulation schemes supported by the 802.16 standard. It uses an outer Reed–Solomon block code in combination with convolution code in the downlink for OFDM-PHY. Turbo coding is an optional code used for this standard to improve the coverage and /or capacity of the system at the cost of increasing complexity and decoding. A total of 52 combinations of modulation and coding schemes are defined in WiMAX as burst profiles. In the downlink and uplink, Binary Phase Shift Keying (BPSK), Quarternary PSK (QPSK), 16-Quadrature Amplitude Modulation (QAM) and 64-QAM are used. 64-QAM is optional in the uplink.

Table 16.2 *Modulation and coding schemes for 802.16d (Ref. [6]), © IEEE 2005*

Rate ID	Modulation rate	Coding	Information bit/symbol	Information bits/ OFDM symbol	Peak data rate in 5 MHz (Mb/s)
0	BPSK	1/2	0.5	88	1.89
1	QPSK	1/2	1	184	3.95
2	QPSK	3/4	1.5	280	6.00
3	16QAM	1/2	2	376	8.06
4	16QAM	3/4	3	568	12.18
5	64QAM	2/3	4	760	16.30
6	64QAM	3/4	4.5	856	18.36

16.5.3 PHY Layer Frame Structure and Access Method

In order to allow for flexible spectrum usage, both Time Division Duplex (TDD) and Frequency Division Duplex (FDD) configurations are supported. TDD uplink PHY is based on a combination of TDMA and DAMA (Demand Assigned Multiple Access) but the downlink channel is based on TDM. In case of FDD, needed to support half-duplex FDD SSs, provision is also made for a TDMA portion of the downlink. This PHY specification operates in a framed format. The system uses a frame size of 0.5, 1, or 2 ms for data transmission. Within each frame, there is a downlink subframe and an uplink subframe.

The downlink subframe begins with necessary information for frame synchronisation and control. These subframes are composed of transmission bursts, which carry medium access control (MAC) information and user data. Each transmission burst, corresponding to a particular SS, is separated from each other by a preamble field and contains several MAC Protocol Data Units (PDUs). The BS determines the length of uplink and downlink subframes dynamically, and the synchronisation between SSs and BSs is done based on a pre-defined time slot.

In FDD operation, the uplink and downlink channels are on **separate frequencies** and allow the system to simultaneously support full-duplex SSs (which can transmit and receive simultaneously) and half-duplex SSs (which do not communicate simultaneously). In case of TDD, the uplink and downlink transmissions share the same frequency but are separated in time. A TDD frame also has a fixed duration and contains one downlink and one uplink subframe. The frame is divided into an integer number of PSs (Packet Services), which help to partition the bandwidth easily [3]. The TDD framing is adaptive in that the link capacity allocated to the downlink versus the uplink may vary. In the TDD case, the downlink subframe comes first, followed by the uplink subframe. In the FDD case, uplink subframe transmissions occur concurrently with the downlink subframe. Figure 16.5 illustrates the TDD frame structure.

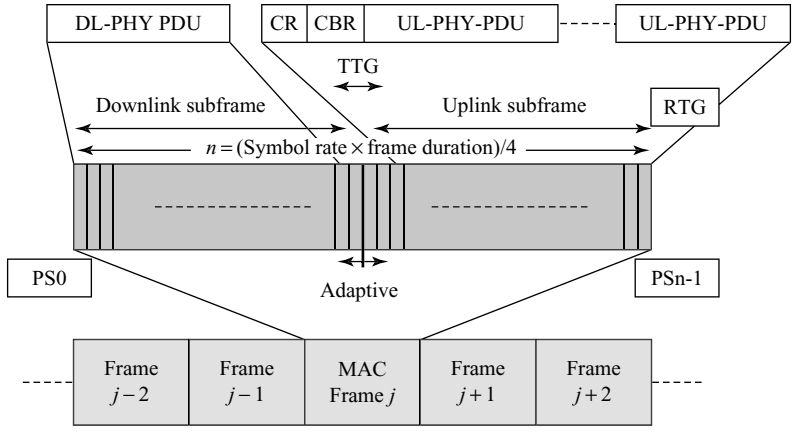


Fig. 16.5 *TDD frame structure (redrawn with the help of [3] ©IEEE copyright 2004)*

There is the TTG (Transmit/Receive Transition Gap) gap between the downlink burst and the subsequent uplink burst. This gap allows times for the BS to switch from the transmit to the receive mode and SSs to switch from the receive to the transmit mode. This gap is an integer number of PS durations and starts on a PS boundary. Similarly, the RTG (Receive/Transmit Transition Gap) is a gap between the uplink burst and the subsequent downlink burst. This gap allows times for the BS to switch from the receive to the transmit mode and the SS to the switch from the transmit to the receive mode [3].

16.5.4 Downlink PHY

The downlink channel is TDM, with the information for each SS multiplexed onto a single stream of data (a burst of MAC Protocol Data Units (PDUs)) and received by all SSs within the same sector. Since the transmission is broadcast, all SSs listen to the data transmitted by the BS. However, an SS is only required to process PDUs that are addressed to itself or that are explicitly intended for all the SSs. The TDD downlink subframe begins with a Frame Start Preamble used by the PHY for synchronization and equalization; and the frame control section, containing DL-MAP, follows the UL-MAP. Fig. 16.6 shows the TDD downlink subframe structure. The DL-MAP message defines the usage of the downlink intervals for a burst mode PHY. The UL-MAP defines the uplink usage in terms of the offset of the burst relative to the allocation start time (units PHY-specific) [3].

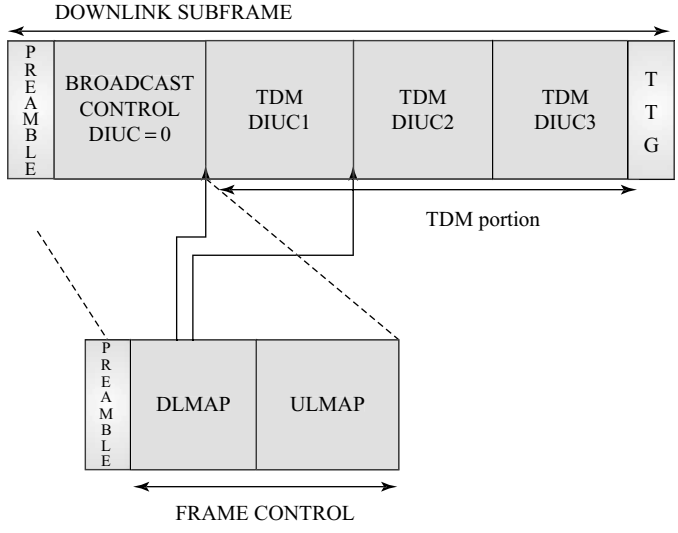


Fig. 16.6 *TDD downlink subframe structure ©IEEE copyright 2004*

The BS at the beginning of each downlink subframe transmits both maps for both FDD and TDD modes. Based on measurements at the physical layer, any SS adapts over time the Interval Usage Code (IUC) in use, that is, the modulation, rate, and Forward Error Correction (FEC) scheme, for both downlink IUC (DIUC) and uplink IUC (UIUC) transmissions. Each SS receives and decodes the control information of the downlink (DL-MAP) and looks for MAC headers indicating data for that SS in the remainder of the downlink subframe. Each SS also learns the boundaries of its allocation within the current uplink subframe by decoding the UL-MAP message. On the other hand, the DL-MAP message contains the timetable of the downlink grants in the forthcoming downlink subframe. More specifically, downlink grants directed to SSs with the same DIUC (Downlink Interval Usage Code) are advertised by the DL-MAP as a single burst.

In the FDD case, the downlink subframe begins with a frame start preamble followed by a frame control section (DL-MAP, UL-MAP) and a TDM portion organized into bursts transmitted in decreasing order of burst profile robustness. This TDM portion of the downlink subframe contains data transmitted to one or more of the following—full-duplex SSs or half-duplex SSs. The FDD downlink subframe continues with a TDMA portion used to transmit data to any half-duplex SSs scheduled to transmit earlier in the frame than when they receive (Fig. 16.7). This allows an individual SS to decode a specific portion of the downlink without the need to decode the entire downlink subframe. In the TDMA portion, each burst begins with the downlink TDMA burst preamble for phase resynchronization. Both FDD and TDD use a burst transmission format whose framing mechanism supports adaptive burst profiling in which transmission parameters, including the modulation and coding schemes, may be adjusted individually to each SS on a frame-by-frame basis.

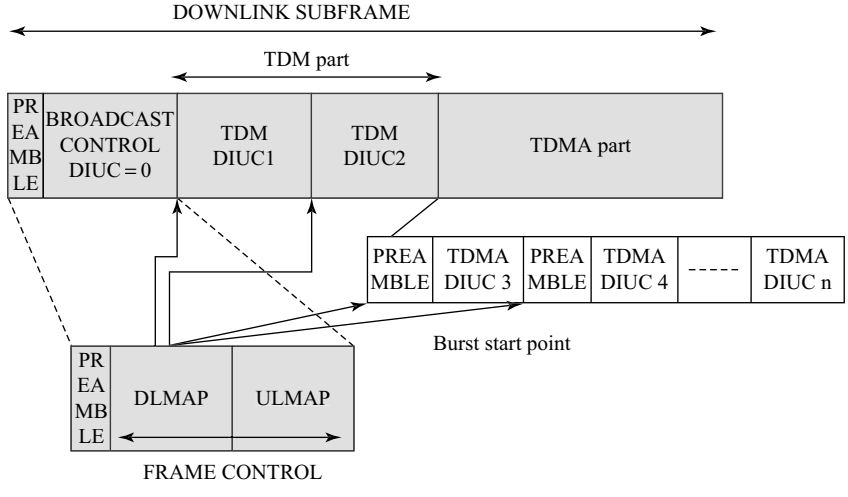


Fig. 16.7 *FDD downlink subframe structure [3], ©IEEE copyright 2004*

The frame control section is the first portion of the downlink frame used for control information destined for all SSs. This control information shall not be encrypted. The information transmitted in this section always uses the well-known downlink burst profile with DIUC = 0. The frame control section shall contain a DL-MAP message for the channel followed by one UL-MAP message for each associated uplink channel. In addition, it may contain DCD (Downlink Channel Descriptor—a MAC message that describes the PHY characteristics of a downlink channel) and UCD (Uplink Channel Descriptor—a medium access control message that describes the PHY characteristics of an uplink) messages following the last UL-MAP message. The BS transmits DCD and UCD at a periodic interval to define the characteristics of downlink physical channel and uplink physical channel.

16.5.5 Uplink PHY

The uplink subframe (Fig. 16.8) is used by the SS to transmit to the BS. Each SS learns the boundaries of its allocation within the current uplink subframe by decoding the UL-MAP message. The UL-MAP contains the information element (IE), which includes the transmission opportunities, i.e., the time slots in which the SS can transmit during the uplink subframe. After receiving the UL-MAP, each SS will transmit data in the predefined time slots as indicated in the IE. The uplink PHY is based on a combination of TDMA and DAMA. In particular, the uplink channel is divided into a number of time slots. The number of slots assigned for various uses (registration, contention, guard, or user traffic) is controlled by the MAC in the BS and may vary over time for optimal performance.

The SS may transmit three classes of bursts during the uplink subframe:

1. Those which are transmitted in contention opportunities reserved for initial ranging.
2. Those which are transmitted in contention opportunities defined by request intervals reserved for response to multicast and broadcast polls.
3. Those which are transmitted in intervals defined by data grant IEs specifically allocated to an individual.

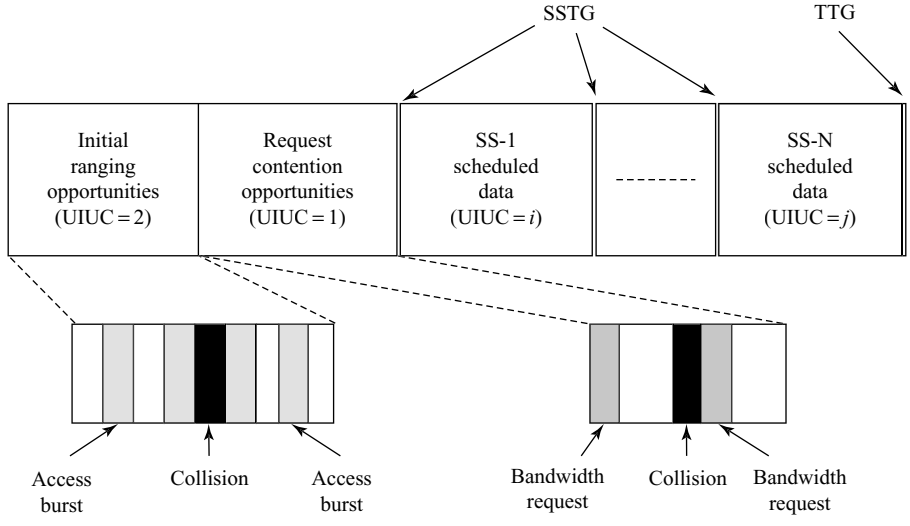


Fig. 16.8 *Uplink subframe structure ©IEEE copyright 2004 [3]*

The bursts may occur in any order and any quantity limited by the number of available packet services and PSs within the frame, at the discretion of the BS uplink scheduler as indicated by the UL-MAP in the frame control section (part of the downlink subframe). The bandwidth allocated for initial ranging and request contention opportunities may be grouped together and is always used with the uplink burst profiles specified for initial ranging intervals (UIUC = 2) and request intervals (UIUC = 1), respectively [3]. The SS groups the remaining transmission slots. During its scheduled bandwidth, an SS transmits with the burst profile specified by the BS. The SSTGs (Subscriber Station Transition Gap) separate the transmissions of the various SSs during the uplink subframe. The gap allows for ramping down of the previous burst, followed by a preamble allowing the BS to synchronise to the new SS. The preamble and gap lengths are broadcast periodically in the UCD message.

Different types of Information Elements (IEs) are supported by SSs. The BS may use any of these IEs when creating a UL-MAP message.

1. **Request IE:** Via the request IE, the BS specifies an uplink interval in which requests may be made for bandwidth for uplink data transmission. If broadcast or multicast, this is an invitation for SSs to contend for requests. If unicast, this is an invitation for a particular SS to request bandwidth.
2. **Initial Ranging IE:** Via the initial ranging IE, the BS specifies an interval in which new stations may join the network by contention.
3. **Data Grant Burst Type IEs:** The Data Grant Burst Type IEs provide an opportunity for an SS to transmit one or more uplink PDUs. These IEs are issued either in response to a request from a station, or because of an administrative policy, such as unicast polling, providing some amount of bandwidth to a particular station.
4. **End of map IE:** An end of map IE terminates all actual allocations in the IE list. It is used to determine the length of the last interval.
5. **Gap IE:** The Gap IE indicates pauses in uplink transmissions. An SS shall not transmit during a Gap IE.

16.6 IEEE 802.16 MAC LAYER OVERVIEW

The MAC is an adaptable and flexible layer, which supports several multiplexing and duplexing schemes. The MAC layer is connection oriented, regardless of the upper layer protocols (connectionless or connection-oriented). The primary task of the WiMAX MAC layer is to provide an interface between the PHY layer and the higher layers. In general, the MAC layer takes packets from the upper layer called the MAC Service Data Unit (MSDUs) and organises them into MAC Protocol Data Unit (MPDUs) for transmission over the air. The IEEE 802.16 fixed and mobile WiMAX standards define a MAC convergence sublayer that can interface with a variety of higher-layer protocols. The MAC comprises three sublayers.

1. Service-Specific Convergence Sublayer (CS).
2. MAC Common Part Sublayer (CPS).
3. Security Sublayer.

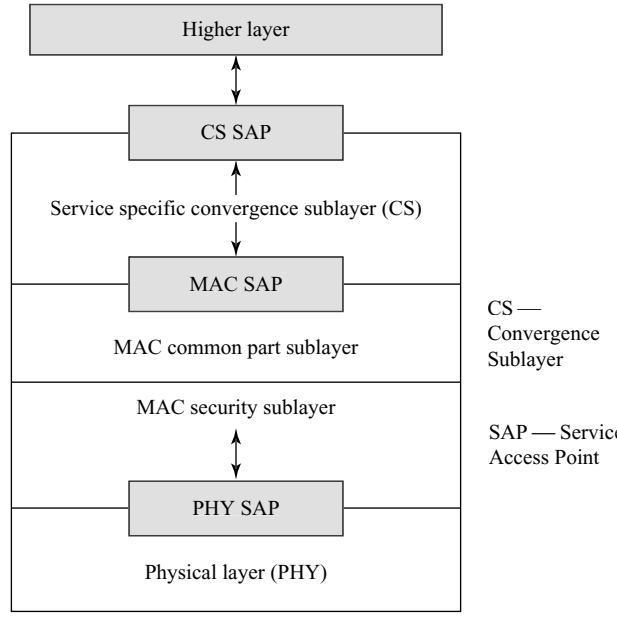


Fig. 16.9 *WiMAX MAC protocol layer*

The general MAC protocol layer architecture is shown in Fig. 16.9.

16.6.1 Service-specific Convergence Sublayer (CS)

The service-specific convergence sublayer provides any transformation or mapping of external network data, received through the CS service access point (SAP), into MAC SDUs received by the MAC common part sublayer through the MAC SAP [3].

It performs the following functionality:

1. Accepting higher-layer PDUs from the higher layer.
2. Performing classification of higher-layer PDUs.
3. Processing the higher-layer PDUs (if required) based on the classification.

4. Delivering CS PDUs to the appropriate MAC SAP.
5. Receiving CS PDUs from the peer entity.

Currently, two CS specifications are provided—the asynchronous transfer mode (ATM) CS and the packet CS.

ATM Convergence Sublayer It resides on top of the IEEE 802.16 MAC common part sublayer. The ATM CS is responsible for accepting ATM cells from the ATM layer and delivering CS PDU to the appropriate MAC-CPS service access point.

The ATM CS performs the following functions:

1. Accepting ATM cells from the ATM layer.
2. Classifying the ATM cells.
3. (If required) processing the ATM cells based on the classification.
4. Delivering the processed ATM cells, i.e., CS PDUs to the appropriate MAC-CPS SAP.

Packet Convergence Sublayer The packet convergence sublayer resides on top of the MAC common part sublayer. The packet CS is responsible for accepting variable-length packets from its upper layers and delivering CS protocol data units to the appropriate MAC-CPS service access point. This packet CS is responsible for transport of all packet-based protocol such as IP, PPP and Ethernet.

The packet CS performs the following functionalities:

1. Classification of the higher-layer protocol PDU into the appropriate connection.
2. Suppression of payload header information (optional).
3. Delivery of the resulting CS PDU to the MAC SAP associated with the service flow for transport to the peer MAC SAP.
4. Receipt of the CS PDU from the peer MAC SAP.
5. Rebuilding of any suppressed payload header information (optional).

The sending CS is responsible for delivering the MAC SDU to the MAC SAP.

MAC SDU Format Once classified and associated with a specific MAC connection, higher-layer PDUs (packet PDUs) shall be encapsulated in the MAC SDU format. The 8-bit Payload Header Suppression Index (PHSI) field needs to be present when a payload header suppression rule has been defined for the associated connection. The MAC SDU format is depicted in Fig. 16.10.

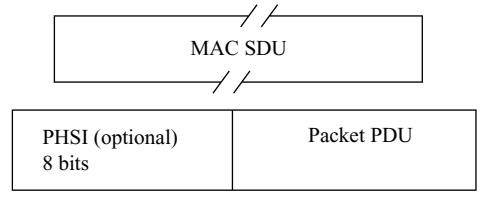


Fig. 16.10 *MAC SDU format ©IEEE copyright 2004 [3]*

Classification Through the classification process, a MAC SDU is mapped onto a particular connection for transmission between MAC peers. This process facilitates the delivery of MAC SDUs with the appropriate Quality of Service (QoS) constraints.

A classifier is a set of matching criteria applied to each packet entering the IEEE Std, 802.16 networks. It consists of some protocol-specific packet matching criteria such as destination IP address, a classifier priority, and a reference to a CID (classifier ID). A packet is delivered to the MAC SAP if it matches the specified packet-matching criteria for delivery on the connection defined by the CID. The service flow characteristics of the connection provide the QoS for that packet. Several classifiers may all refer to the same service flow. The classifier priority is used for ordering the application of classifiers to packets. The priority need not be unique.

1. **Service flow (SF):** A unidirectional flow of medium access control service data units (SDUs) on a connection that is provided a particular quality of service.
   The packet classification table contains the following fields related to different classifiers:
2. **Priority:** Determines the search order for the table. Higher priority classifiers are searched before lower priority classifiers.

3. **Ethernet/IEEE 802.3 classification parameters:** Zero or more of the Ethernet/IEEE 802.3 classification parameters (destination MAC Address, source MAC address, ether type/SAP).
4. **IEEE 802.1P/Q parameters:** Zero or more of the IEEE 802.1P/Q classification parameters (802.1P priority range, 802.1Q VLAN ID).
5. **IPv4/IPv6 classification parameters:** Zero or more of the IP classification parameters (IPv4 TOS range/mask, IPv4 protocol, IPv4 source address/mask, IPv4 destination address/mask).
6. **TCP/UDP classification parameters:** TCP/UDP Source Port Start, TCP/UDP Source Port End, TCP/UDP Destination Port Start, TCP/UDP Destination Port End).
7. **Service flow Identifier:** Identifier of a specific flow to which this packet is to be directed.

Classifiers can be added to the table either via management operations (configuration file, registration, and SNMP), or via dynamic operations (dynamic signaling, MAC-CPS SAP).

BS applies downlink classifiers to transmitting packets and uplink classifiers are applied at the SS. Figure 16.11 specifies that when a packet comes from an upper layer, it is compared against a set of classifiers. The matching classifier for the packet identifies the corresponding service flow via the Service Flow ID (SFID). In the case where more than one classifier matches the packet, the highest priority classifier is chosen.

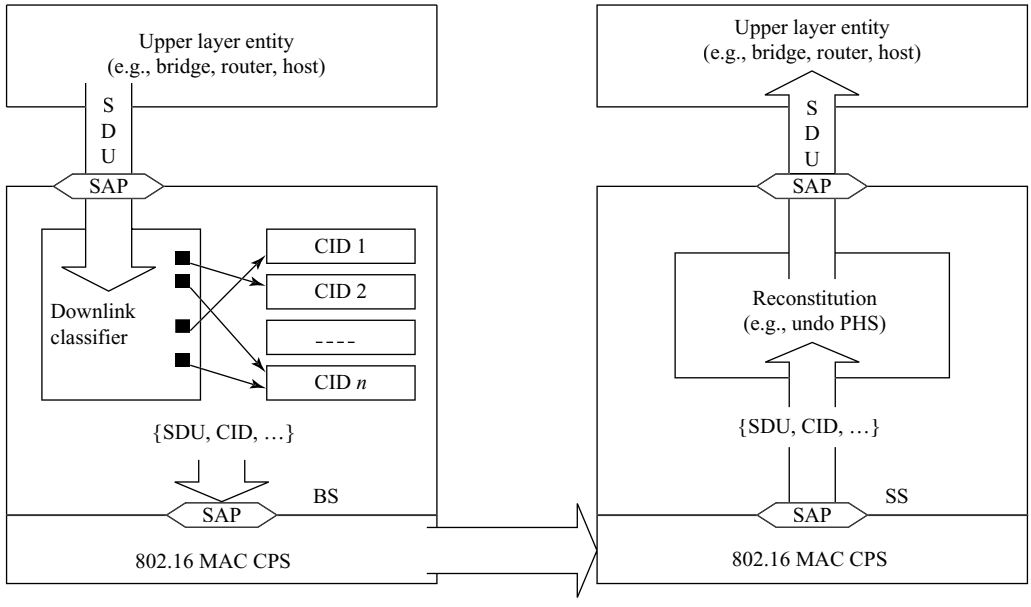


Fig. 16.11 *Classification and CID mapping (BS to SS) [3], ©IEEE copyright 2004*

The classifier matching a packet may be associated with a Payload Header Suppression Rule (PHS rule). When a service flow is deleted, all classifiers and any associated PHS rules referencing it shall also be deleted. It is possible for a packet to fail to match the set of defined classifiers. In this case, the packet CS service type may either associate the packet with its assigned default CID or discard the packet.

Payload Header Suppression: PHS The payload header suppression field is a string of bytes representing the header portion of a Protocol Data Unit (PDU) in which one or more bytes are to be suppressed. Implementation of PHS is optional. In PHS, the repetitive part of the payload headers of the higher layer is suppressed in the MAC SDU by the sending entity and restored by the receiving entity. On the uplink, the sending entity is the SS and the receiving entity is the BS. On the downlink, the sending entity is the BS and the receiving entity is the SS. If PHS is enabled at MAC connection, each MAC SDU is prefixed with a PHSI, which references the Payload Header Suppression Field (PHSF).

The sending entity uses classifiers, which uniquely maps packets to its associated PHS rule. The receiving entity uses the CID and the PHSI to restore the PHSF. In the sending end once a PHSF has been assigned to a PHSI, it will be unchanged. To change the value of a PHSF on a service flow, a new PHS rule shall be defined. When a classifier is deleted, any associated PHS rule shall also be deleted. It is the responsibility of the higher-layer service entity to generate a PHS rule that uniquely identifies the suppressed header within the service flow.

The **PHS has a Payload Header Suppression Valid (PHSV)** option to verify or not verify all bytes that are to be suppressed in the payload header before suppressing. The PHS also a Payload Header Suppression Mask (PHSM) option to allow selection of bytes that are not to be suppressed. Either the sending or the receiving entity shall specify the PHSF and the Payload Header Suppression Size (PHSS).

Payload Header Suppression Size (PHSS) It is the length of the suppressed field in bytes. This value is equivalent to the number of bytes in the Payload Header Suppression Field (PHSF) and also the number of valid bits in the Payload Header Suppression Mask (PHSM).

16.6.2 MAC Common Part Sublayer (MAC CPS)

MAC CPS is a core layer of the IEEE 802.16 standard. This provides core MAC functionality like bandwidth allocation, connection establishment and connection maintenance [3]. They are of different types of connections established by the MAC. The first one is signaling connections, which follow specific rules and can be for different purposes and the second one is data connections, which are defined with a service flow, that is quality of service parameters.

The main operations of the MAC layer performed by this layer are addressing and quality of service. As the IEEE 802.16 standard has been designed to fully support multiple network layer protocols, either ATM or packets (IP, Ethernet, etc.), so, Quality of Service (QoS) is a key feature of the standard to support multiple types of traffic that are carried through the network. Achieving QoS for the whole system is made possible by the standard through the combination of multiple features as outlined under:

1. Connection allocation, modification, and deletion on either subscriber or base-station request.
2. Fragmentation of MAC Service Data Units (MSDUs) in order to fit in allocated time slots; this parameter is set on a per-connection basis.
3. Packing of MAC SDUs in one allocated time slot; this parameter is set on a per-connection basis.
4. Dynamic setting of the uplink map.
5. Dynamic setting of the downlink map.

MAC CPS is not required to understand the format of any information from the CS payload. It receives data from the various CSs, through the MAC SAP, classified to particular MAC connections. The actual media-access methods were provided by the point-to-multipoint architecture and mesh architecture.

MAC PDU Format Each PDU shall begin with a fixed-length MAC header (i.e., 6 bytes). Payload of the MAC PDU may follow the header. If payload is present, it shall consist of zero or more subheaders and zero or more MAC SDUs and/or fragments thereof. The payload information may vary in length, so that a MAC PDU may represents a variable number of bytes. A MAC PDU may contain a CRC also. Figure 16.12 represents a MAC PDU format.

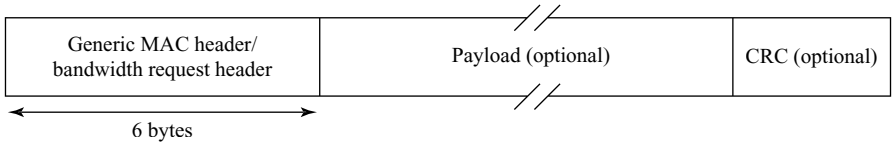


Fig. 16.12 *MAC PDU format [3] ©IEEE copyright 2004*

The MAC header may be a 'Generic MAC Header' or a 'Bandwidth Request Header'. A Generic MAC Header (GMH) is used to send either a MAC management message or CS data, and each GMH is encoded. The generic header format is illustrated in Fig. 16.13.

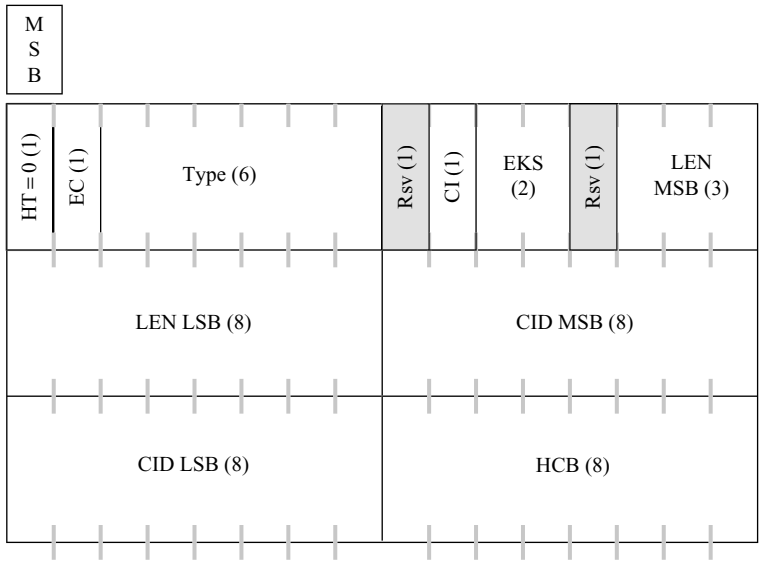


Fig. 16.13 *Generic MAC header format © IEEE copyright 2006*

Description of the Header Fields

1. **HT (Header Type):** For a generic MAC header this bit should be set to zero. It indicates whether the header is a generic MAC header or a bandwidth request header.
2. **EC (Encryption Control):** This bit specifies whether the payload is encrypted or not.
   If EC = 0 then not Encrypted
   Else Encrypted
3. **Type:** This field indicates the subheaders and special payload types present in the message payload. Each bit indicates the following functionality.

   (a) Bit 0 is set when a grant management subheader is present in the payload.
   (b) Bit 1 is set when a packing subheader is present in the payload.
   (c) Bit 2 is set when a fragmentation subheader is present in the payload.
   (d) Bit 3 is set when the fragmentation or packing headers are extended.
   (e) Bit 4 is set when the frame contains an ARQ feedback payload.
   (f) Bit 5 is set when a mesh subheader is present.

4. **CI (CRC Indicator):** If 0 then no CRC is included. If 1 then CRC is included in the PDU by appending it to the PDU payload after encryption.
5. **EKS (Encryption Key Sequence):** The index of the Traffic Encryption Key (TEK) and initialization vector used to encrypt the payload. This field is only meaningful if the EC field is set to 1.
6. **LEN (Length):** The length in bytes of the MAC PDU includes the MAC header and the CRC if present.
7. **CID:** Connection identifier
8. **HCS (Header Check Sequence):** An 8-bit field used to detect errors in the header. The transmitter shall calculate the HCS value for the first five bytes of the cell header, and insert the result into the HCS

field (the last byte of the MAC header). It shall be the remainder of the division (Modulo 2) by the generator polynomial g (D = D8 + D2 + D + 1) of the polynomial D8 multiplied by the content of the header excluding the HCS field. (Example: [HT EC Type] = 0 × 80, BR = 0 × AAAA, CID = 0 × 0F0F; HCS should then be set to 0 × D5).

Bandwidth Request Header The bandwidth request PDU shall consist of a bandwidth request header alone and shall not contain a payload. The bandwidth request header is given in Fig. 16.14. Every header is encoded.

The bandwidth request shall have the following properties [3]:

1. The length of the header is always 6 bytes.
2. The EC field shall be set to 0 to indicate no encryption.
3. The Bandwidth Request (BR) field shall indicate the number of bytes of uplink bandwidth requested by the SS. The bandwidth request is for the CID. The request shall not include any PHY overhead.
4. The CID shall indicate the connection for which uplink bandwidth is requested.
5. The allowed types for bandwidth requests are '000' for incremental and '001' for aggregate.

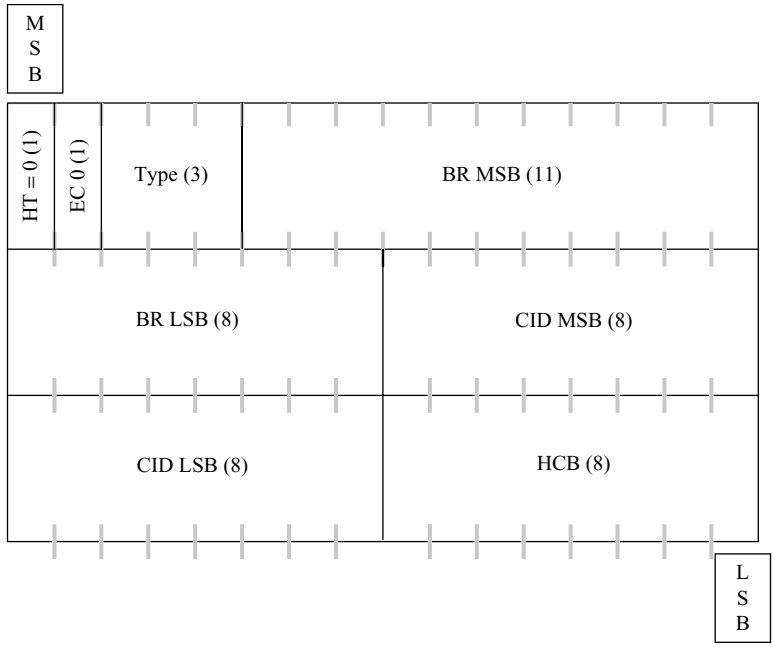


Fig. 16.14 *Bandwidth request header format © IEEE copyright 2006*

MAC Subheaders There are five types of MAC subheaders—grant management subheader, fragmentation subheader, packing subheader, mesh subheader and FAST-FEEDBACK allocation subheader.

An SS uses the **grant management subheader**. It conveys bandwidth management needed to its BS. This subheader is encoded differently based upon the type of uplink scheduling service (TDD/FDD) for the connection. The capability of a grant management subheader at both the BS and SS is optional.

The **fragmentation subheader** contains information that indicates the presence and orientation in the payload of any fragments of SDUs.

The **packing subheader** is used to indicate the packing of multiple SDUs into a single PDU. When packing variable-length MAC SDUs, the MAC precedes each one with a packing subheader.

The **mesh subheader** is used in the mesh mode. The mesh subheader always follows the generic MAC header.

The SS needs to send a feedback message for the connection associated with the CID value uses the FAST-FEEDBACK allocation subheader. This subheader shall always be the last per-PDU subheader. Out of these 5 subheaders, mesh, fragmentation, FAST-FEEDBACK allocation, and grant management subheaders are per-PDU subheaders. It means that in a single MAC PDU, only one of each subheader may be included. The only per-SDU subheader is the packing subheader. It means that in each SDU coming from the CS through MAC SAP, this subheader may be included. The per-PDU subheaders may be inserted in MAC PDUs immediately following the generic MAC header. If both the fragmentation subheader and grant management subheader are indicated, the grant management subheader shall come first. If the mesh subheader is indicated, it shall precede all other subheaders. The packing and fragmentation subheaders are mutually exclusive and shall not both be present within the same MAC PDU.

MAC Management Messages All MAC management messages are carried in the payload part of the MAC PDU. All MAC management messages begin with a management message type field of 8 bits and may contain additional fields of variable length. The MAC management message format is given in Fig. 16.15.

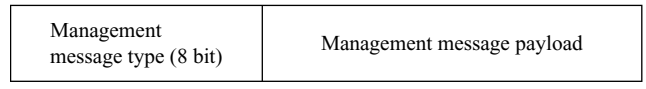


Fig. 16.15 *MAC management message format*

For the basic, broadcast, and initial ranging connections, MAC management messages shall neither be fragmented nor packed. These messages on the primary management connection may be packed and/or fragmented. MAC management messages cannot be carried on Transport Connections.

Some important MAC management messages for PMP architecture are described below.

1. **Downlink Channel Descriptor (DCD) message:** A DCD is transmitted by the BS at a periodic interval to define the characteristics of a downlink physical channel.
2. **Downlink map (DL-MAP) message:** The DL-MAP message defines the access to the downlink information. The DL-MAP message contains the timetable of the downlink grants in the forthcoming downlink subframe.
3. **Uplink Channel Descriptor (UCD) message:** A UCD is transmitted by the BS at a periodic interval to define the characteristics of an uplink physical channel.
4. **Uplink map (UL-MAP) message:** The UL-MAP message allocates access to the uplink channel. Each SS learns the boundaries of its allocation within the current uplink subframe by decoding the UL-MAP message.
5. **Ranging request (RNG-REQ) message:** An RNG-REQ is transmitted by the SS at initialization, to determine network delay periodically and to request power and/or downlink burst profile change.
6. **Ranging response (RNG-RSP) message:** The adjustments to the SS's transmit time advance, as well as power adjustments, are returned to the SS in ranging response (RNG-RSP) messages. For ongoing ranging and power adjustments, the BS may transmit unsolicited RNG-RSP messages commanding the SS to adjust its power or timing.

   The SS transmits a **Registration Request** (REG-REQ) message at initialization.

   A **Registration Response** (REG-RSP) message is transmitted by the BS in response to a received REG-REQ.

   A DSA (**Dynamic Service Addition**)-REQ message is sent by an SS or BS to create a new service flow. A DSA-REQ message does not contain parameters for more than one service flow. There are two types of DSA-REQ messages—SS-initiated DSA, and BS-initiated DSA.

   A DSA-RSP message (DSA Response) is generated in response to a received DSA-REQ. There are two types of DSA-RSP messages—SS-initiated DSA-RSP, and BS-initiated DSA-RSP.

7. **DSA-ACK message:** A DSA-ACK is generated in response to a received DSA-RSP.
8. **DSC (dynamic service change) Request (DSC-REQ) message:** A DSC-REQ is sent by an SS or BS to dynamically change the parameters of an existing service flow. But a DSC-REQ message shall not carry parameters for more than one service flow.
9. **DSC Response (DSC-RSP) message:** A DSC-RSP is generated in response to a received DSC-REQ.
10. **DSC Acknowledge (DSC-ACK) message:** A DSC-ACK is generated in response to a received DSC-RSP.
11. **DSD-REQ message:** A DSD-REQ is sent by an SS or BS to delete an existing service flow.
12. **DSD-RSP message:** A DSD-RSP is generated in response to a received DSD-REQ.
13. **Multicast Polling Assignment Request (MCA-REQ) message:** The MCA-REQ message is sent to an SS to assign it to or remove it from a multicast-polling group.
14. **Multicast Polling Assignment Response (MCA-RSP) message:** The SS in response to a MCA-REQ sends the MCA-RSP.
15. **Downlink Burst Profile Change Request (DBPC-REQ) message:** The DBPC-REQ message is sent by the SS to the BS on the SS's Basic CID to request a change of the downlink burst profile used by the BS to transport data to the SS. Note that a change of downlink burst profile may also be requested by means of an RNG-REQ.
16. **Downlink Burst Profile Change Response (DBPC-RSP) message:** The DBPC-RSP message is transmitted by the BS on the SS's Basic CID in response to a DBPC-REQ message from the SS.
17. **Reset Command (RES-CMD) message:** The RES-CMD message is transmitted by the BS on an SS's basic CID to force the SS to reset itself, reinitialise its MAC, and repeat initial system access. This message may be used if an SS is unresponsive to the BS or if the BS detects continued abnormalities in the uplink transmission from the SS.

Creation of Protocol Data Unit For proper utilization of air link resources, the 802.16 MAC layer performs both fragmentation and packing of MAC SDUs.

1. **Fragmentation:** Fragmentation is the process by which a MAC SDU is divided into one or more MAC PDUs that are transmitted independently. This technique provides an efficient use of available bandwidth relative to the QoS requirements of a connection's service flow. The authority to fragment traffic on a connection is defined when the MAC SAP creates the connection. To indicate the fragmentation in a MAC PDU, a fragment subheader (FSH) is included at the start of the payload, as shown in Fig. 16.16. A MAC PDU consists of a fixed-length MAC header, a variable length payload, and an optional cyclic redundancy check (CRC) [8].

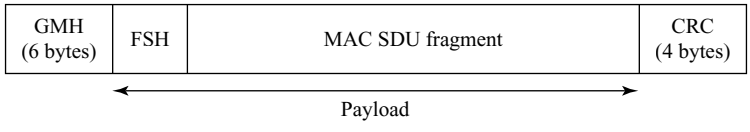


Fig. 16.16 *MAC PDU with fragment sub-header, ©IEEE copyright 2002*

The fragment control (FC) bits of a fragmentation subheader is indicated as follows:

(a) 10 - fragment is the first fragment of an MAC SDU
(b) 01 - fragment is the last fragment of an MAC SDU
(c) 11 - fragment is somewhere in the middle of an MAC SDU
(d) 00 - MAC SDU is not fragmented

The fragment sequence number (FSN) of the fragmentation subheader increases by 1 for each fragment of a MAC SDU so the receiver can reassemble fragments appropriately.

2. **Packing:** Multiple MAC SDUs or multiple MAC SDU fragments can be packet into a single MAC SDU. This is sometimes referred as MAC-level packet aggregation. A MAC PDU can contain multiple packing subheaders, each followed by either a MAC SDU or a fragment of a MAC SDU. Packing supports both the connections that carry fixed length packets and on those that carry variable length packets.
3. **Packing fixed-length SDUs:** In case of packing fixed length SDU, the MAC simply needs to know the size of a fixed length SDU that will have to be specified at the time of connection establishment. Depending on the size specified in the length field of the generic MAC header, the receiver can identify the number of SDUs packed in a single MAC PDU. If the MAC SDU size is n bytes, the receiving side can unpack simply by knowing that the length field in the MAC header will be $n \times k + j$, where k is the number of MAC SDUs packed into the MAC PDU and j is the size of the MAC header and any MAC subheaders. Figure 16.17 illustrates the MAC SDUs packing for fixed length.

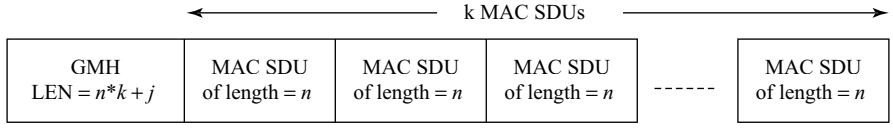


Fig. 16.17 *Packing fixed-length MAC SDUS into a single MAC PDU, ©IEEE copyright 2004*

4. **Packing variable-length SDUs:** In the variable-length MAC SDU case, the MAC attaches a packing subheader to each MAC SDU. The packing subheader starts with a single length extension (LE) bit, which is set to 0 if the length field in the packing subheader is 7 bits or set to 1 if it is 15 bits. If the packing subheader length is 1 byte then it allows small SDUs of less than 128 bytes, and if it is 2 bytes then it allows small SDUs of more than 128 bytes. The packing subheaders with LE = 0 and LE = 1 may both be present in the same MAC PDU.

   Both unfragmented and fragmented MAC SDUs may be present in the same MAC PDU. If more than one MAC SDU is packed into the MAC PDU, the type field in the MAC header indicates the presence of Packing Subheaders (PSHs). If a MAC SDU does not fit into the remainder of an MAC PDU, then it can be allocated to occupy the remainder of the current MAC PDU, and the rest will be send in the subsequent MAC PDUs. A MAC PDU format, which consists of variable-length MAC SDUs, is given in Fig. 16.18.

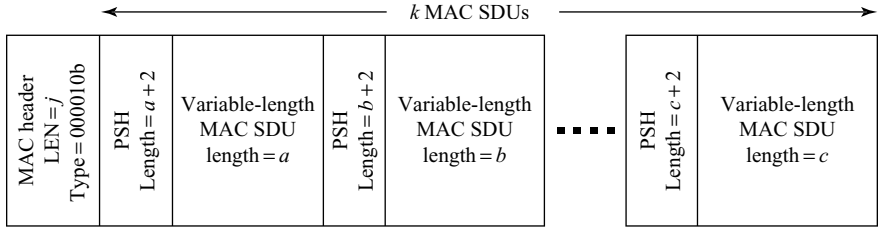


Fig. 16.18 *Packing variable-length MAC SDUs into a single MAC PDU, ©IEEE copyright 2004*

Automatic Repeat Request (ARQ) ARQ processing is the process of retransmitting MAC SDU blocks (ARQ blocks) that have been lost or damaged. The 802.16 MAC uses a simple sliding window-based approach, where the transmitter can transmit up to a negotiated number of blocks without receiving an acknowledgement. The receiver sends an acknowledgement or a negative acknowledgement messages to indicate to the transmitter as to which SDU blocks have successfully been received and which have been lost. The transmitter retransmits blocks that were lost and moves the sliding window forward when the SDU blocks acknowledged have been received.

16.6.3 Security Sublayer

The IEEE 802.16 security is implemented as a security sulayer at the bottom part of the MAC layer. The security sublayer provides the security features across the fixed broadband wireless network by encrypting connections between SS and BS. Depending upon the different service flows used for different connections the BS protects against unauthorized access to these data transport services through different encryption techniques by using a client/server key-management protocol. In this client/server key-management protocol, a BS acts as a server and each SS is represented as a client. Here, the BS's responsibility is to distribute the key material to each client.

16.7 IEEE 802.16 SCHEDULING SERVICES

The IEEE 802.16 MAC, defined as connection-oriented, is designed to support different QoS for different types of services. In the IEEE 802.16 standard, both for fixed and mobile WiMAX, the broadband access requirements can be classified into four types according to the scheduling services—Unsolicited Grant Service (UGS), real-time Polling Service (rtPS), non-real time Polling Service (nrtPS) and Best Effort (BE). The current draft adds a new scheduling service called extended real-time Polling Service (ertPS), which combines the efficiency of UGS and rtPS. It allows unsolicited bandwidth grants like UGS, but with a dynamic size like that of rtPS. This yields a service class supporting real-time service flow with variable-size data packets, suitable for Voice-over IP (VoIP) with silence suppression. Providing guaranteed QoS for all types of service flows in the IEEE 802.16 wireless MAN is a real challenging problem.

The following items are taken into account for each active service flow:

1. The scheduling service specified for the service flow.
2. The values assigned to the service flow's QoS parameters.
3. The availability of data for transmission.
4. The capacity of the granted bandwidth.

Scheduling services represent [3] [9] the data handling mechanisms supported by the MAC scheduler for data transport on a connection. Each connection is associated with a single data service. Each data service is associated with a set of QoS parameters that quantify aspects of its behavior. These parameters are managed using the dynamic service addition (DSA) and dynamic service change (DSC) message dialogs. In this section, each type of service is described.

Figure 16.19 shows the QoS architecture of the IEEE 802.16. The blocks drawn with dotted lines in Fig. 16.19 are open or undefined [7]. The BS uplink packet scheduling using bandwidth request sent from SSs to BS determines the information element (IE), which includes the transmission opportunities, i.e., the time slots in which the SSs can transmit during uplink subframe. The uplink packet scheduling is only defined for UGS service flow but for other service flows, it is undefined.

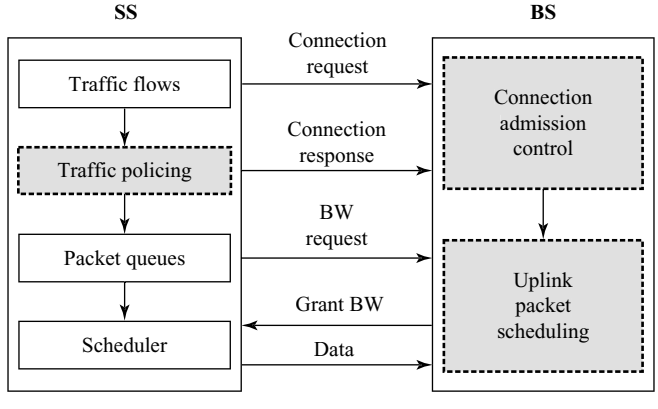


Fig. 16.19 *QoS architecture of IEEE 802.16*

16.7.1 Unsolicited Grant Service (UGS)

UGS is designed to support real-time applications (with strict delay requirements) that generate fixed-size data packets at periodic intervals, such as T1/E1 and VoIP without silence suppression. The service eliminates the overhead and latency of SS requests and assures that grants are available to meet the flow's real-time needs. The BS provides data grant burst IEs to the SS at periodic intervals based upon the maximum sustained traffic rate of the service flow. These applications require constant bandwidth allocation, so that bandwidth requests are not required. The size of these grants shall be sufficient to hold the fixed-length data associated with the service flow (with an associated generic MAC header and a grant management subheader) but may be larger at the discretion of the BS scheduler. In order that this service work correctly, the request/transmission policy setting shall be such that the SS is prohibited from using any contention request opportunities as well as a piggyback request for this connection. The mandatory QoS service flow parameters for this scheduling service are minimum reserved traffic rate, maximum sustained traffic rate, maximum latency, tolerated jitter, and request/transmission policy. Figure 16.20 shows the information flow for UGS [11].

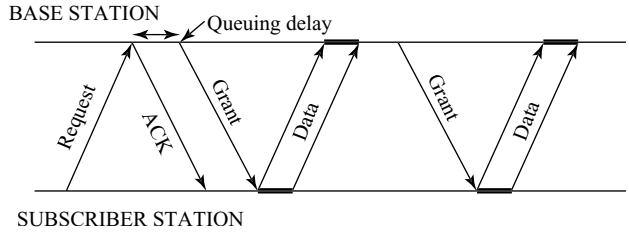


Fig. 16.20 *Information flow of UGS*

16.7.2 Real-time Polling Service (rtPS)

rtPS is designed to support real-time applications (with less stringent delay requirements) that generate variable-size data packets at periodic intervals, such as Moving Pictures Expert Group (MPEG) video and VoIP with silence suppression. The service offers real-time, periodic unicast request opportunities, which meet the flow's real-time needs and allow the SS to specify the size of the desired grant. Since the size of arriving packets with rtPS is not fixed, as it is with UGS-tailored applications, rtPS connections are required to notify the BS of their current bandwidth requirements. This service requires more request overhead than UGS, but supports variable grant sizes for optimum data transport efficiency. In order for this service to work correctly, the request/transmission policy setting shall be such that the SS is prohibited from using any contention request opportunities as well as piggyback requests for that connection. These applications have specific bandwidth requirement, as well as a maximum acceptable latency (tight delay bound). Late packets that miss the deadline are considered useless. The mandatory QoS service flow parameters for this scheduling service are minimum reserved traffic rate, maximum sustained traffic rate, maximum latency, and request/transmission policy.

A fundamental difference between UGS traffic and rtPS traffic is that UGS reserves a fixed portion of upstream bandwidth that can only be used by that service flow. In rtPS, however, if the service flow is inactive for a short period of time, the excess reserved capacity can be reused by other rtPS (or nrtPS and BE) flows. Figure 16.21 shows the information flow for UGS [11].

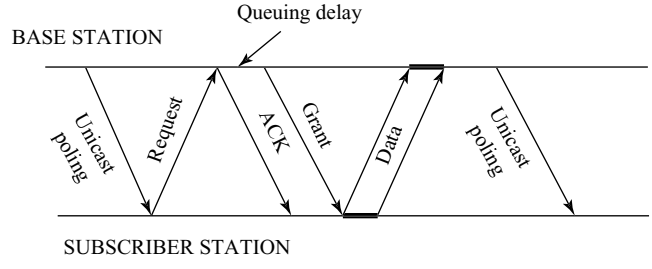


Fig. 16.21 *Information flow of rtPS*

16.7.3 Non-real-time Polling Service (nrtPS)

Unlike UGS and rtPS scheduling services, nrtPS and BE are designed for applications that do not have any specific delay requirement (non-real time application). The nrtPS is designed to support delay-tolerant data streams consisting of variable-sized data packets for which a minimum data rate is required, such as FTP. The nrtPS offers unicast polls on a regular basis, which assures that the service flow receives bandwidth request opportunities even during network congestion. The BS typically polls nrtPS CIDs on an interval less frequently than rtPS. In case the dedicated request opportunities cannot satisfy the flow's bandwidth requirements, the contention request opportunities (broadcast poll) as well as piggyback requests are allowed to be used as well. This results in the SS using contention request opportunities as well as unicast request opportunities and unsolicited data grant burst types. The mandatory QoS service flow parameters for this scheduling service are minimum reserved traffic rate, maximum sustained traffic rate, traffic priority and request/transmission policy.

16.7.4 Best Effort (BE)

The BE service is designed to support data streams for which no minimum service level is required and, therefore, may be handled on a space-available basis. The intent of the BE service is to provide efficient service for best-effort traffic. This service is used for traffic requiring no QoS guarantee. In order for this service to work correctly, the request/transmission policy setting shall be set such that the SS is allowed to use contention request opportunities (broadcast poll) as well as piggyback requests. This results in the SS using contention request opportunities as well as unicast request opportunities and unsolicited data grant burst types. Applications of this kind shares the remaining bandwidth after allocation to the previous three services is completed. The main difference between nrtPS and BE is that nrtPS connections are reserved for a minimum amount of bandwidth (by means of minimum reserved traffic rate). The mandatory QoS service flow parameters for this scheduling service are maximum sustained traffic rate, traffic priority, and request/transmission policy. Figure 16.22 shows the information flow for BE services.

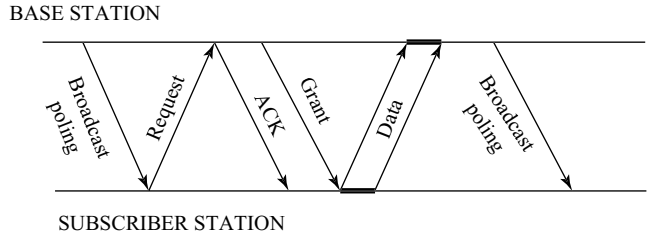


Fig. 16.22 *Information flow of BE*

16.8 BANDWIDTH ALLOCATION AND REQUEST MECHANISMS

During network entry and initialisation, every SS is assigned up to three dedicated CIDs for sending and receiving control messages. These connection pairs are used to allow differentiated levels of QoS to be applied to the different connections carrying MAC management traffic. Apart from incompressible constant bit rate UGS connections, increase or decrease of bandwidth requirements is necessary for all services [3]. When an SS requests for bandwidth connection with a BE scheduling service or rtPS or nrtPS, it sends a message to the BS. The QoS for the connection is established at the connection establishment and is looked up by the BS. In case of PMP mode, there are numerous methods by which the SS can get the bandwidth request message from the BS such as requests, grants, and polling.

16.8.1 Requests

Request is the mechanism which SSs use to indicate to the BS that they need uplink bandwidth allocation. A request may be a standalone bandwidth request header or it may be a piggyback request (optional). The bandwidth request message may be transmitted during any uplink allocation, except during any initial ranging interval. Because the uplink burst profile can change dynamically, all requests for bandwidth shall be made

in terms of the number of bytes needed to carry the MAC header and payload, but not the PHY overhead [3]. Bandwidth requests may be incremental or aggregate. In case of an incremental bandwidth request, the BS shall add the quantity of bandwidth requested to its current perception of the bandwidth needs of the connection. When the BS receives an aggregate bandwidth request, it shall replace its perception of the bandwidth needs of the connection with the quantity of the bandwidth requested. There is a type field in the bandwidth request header used to indicate whether the request is incremental or aggregate. Since piggybacked bandwidth requests do not have a type field, they shall always be incremental. The self-correcting nature of the request/grant protocol requires that SSs periodically use aggregate bandwidth requests. The period may be a function of the QoS of a service and of the link quality. Due to the possibility of collisions, bandwidth requests transmitted in broadcast or multicast request IEs should be aggregate requests. Additional bandwidth request mechanisms include the focused bandwidth requests and CDMA bandwidth requests.

16.8.2 Grants

For an SS, the bandwidth requests a reference individual connection while each bandwidth grant is addressed to the SS's basic CID, and not to individual CIDs. Based on the latest information received from the BS and the status of the request, the SS may decide to perform backoff and request again or to discard the SDU. An SS may use request IEs (information elements) that are broadcast and directed at a multicast-polling group in which it is a member, or directed at its basic CID. In all cases, the request IE burst profile is used, even if the BS is capable of receiving the SS with a more efficient burst profile. For more efficient use of a burst profile, the SS should transmit in an interval defined by a data grant IE directed at its basic CID in a unicast polling in general. In a data grant IE directed at its basic CID, the SS may make bandwidth requests for any of its connections.

16.8.3 Polling

Through the polling process, the BS allocates bandwidths specifically to the SSs for the purpose of making bandwidth requests to individual SSs or to groups of SSs. The allocations are contained as a series of IEs within the UL-MAP instead of explicit messages. Bandwidth is always requested based on CID, and bandwidth is allocated on an SS basis.

Different types of polling are unicast, multicast and broadcast, and PM bit.

1. **Unicast:** SS is polled individually. No explicit message is transmitted to poll the SS. Rather, the SS is allocated in the UL-MAP bandwidth, sufficient to respond with a bandwidth request. If the SS does not need a bandwidth, the allocation is padded. SSs that have an active UGS connection of sufficient bandwidth are not polled individually unless they set the PM bit in the header of a packet on the UGS connection, thus saving bandwidth over polling all SSs individually. Unicast polling would normally be done on a per-SS basis by allocating a data grant IE directed at its basic CID [3].
2. **Multicast and Broadcast:** If insufficient bandwidth is available to individually poll many inactive SSs, some SSs may be polled in multicast groups or a broadcast poll may be issued. Some definite CIDs are reserved for multicast groups and for broadcast messages. As with individual polling, the poll is not an explicit message, but a bandwidth allocation in the UL-MAP with the difference that rather than associating an allocated bandwidth with an SS's basic CID, the allocation is to a multicast or broadcast CID. During a multicast polling or broadcast CID, an SS belonging to the polled group may request bandwidth during any request interval allocated to that CID in the UL-MAP by using a request IE. In order to reduce the likelihood of collision with multicast and broadcast polling, a contention algorithm is applied to select the slot for transmission with the initial bandwidth request only for SS's needing a bandwidth reply. Zero-length bandwidth (BW) requests shall not be used in multicast or broadcast request intervals. The SS shall assume that the transmission has been unsuccessful if no grant has been received in the number of subsequent UL-MAP messages specified by the parameter contention-based reservation timeout.

3. **PM bit:** The PM bit in the grant management subheader is set for the SSs with currently active UGS connections in a MAC packet of the UGS connection which indicate to the BS that they need to be polled in order to request bandwidth for non-UGS connections. Only the SSs with active UGS connections need be individually polled to reduce the bandwidth requirements of individual polling only if the PM bit is set (or if the interval of the UGS is too long to satisfy the QoS of the SS's other connections). Once the BS detects this request for polling, the process for individual polling is used to satisfy the request.

In case of mesh mode, bandwidth allocation and request mechanisms are different from the PMP mode.

16.9 NETWORK ENTRIES AND INITIALISATION

In case of PMP mode, 802.16 systems support the applicable procedures for entering and registering a new SS or a new node to the network [3] [9]. The procedures are divided into the following phases:

1. Scan for downlink channel and establish synchronization with the BS.
2. Obtain transmit parameters (from UCD message).
3. Perform ranging.
4. Negotiate basic capabilities.
5. Authorise SS and perform key exchange.
6. Perform registration.
7. Establish IP connectivity.
8. Establish time of day.
9. Transfer operational parameters.
10. Set-up connections.

Each SS contains the following information when shipped from the manufacturer:

1. A 48-bit universal MAC address assigned during the manufacturing process. This is used to identify the SS to the various provisioning servers during initialization.
2. Security information as defined in Clause 7 (e.g., X.509 certificate) used to authenticate the SS to the security server and authenticate the responses from the security and provisioning servers.

16.9.1 Scanning and Synchronization to the Downlink

The SS shall acquire a downlink channel on initialization or after signal loss. Due to the non-volatile storage in SSs, the last operational parameters are stored, and the SS shall first try to reacquire this downlink channel. If it fails, it shall begin to continuously scan the possible channels of the downlink frequency band of the operation until it finds a valid downlink signal. MAC attempts to acquire channel control parameters for the downlink once the PHY has achieved synchronization, as given by a PHY Indication and then it attempts to acquire the uplink.

16.9.2 Obtain Downlink Parameters

The MAC shall search for the DL-MAP MAC management messages. The SS achieves MAC synchronization once it has received at least one DL-MAP message. An SS MAC remains in synchronization as long as it continues to successfully receive the DL-MAP and DCD messages for its channel.

16.9.3 Obtain Uplink Parameters

After the synchronization process, the SS waits for a UCD message from the BS in order to retrieve a set of transmission parameters for a possible uplink channel. These messages are transmitted periodically from the BS for all available uplink channels and are addressed to the MAC broadcast address. If no uplink channel can be found after a suitable timeout period then the SS shall continue scanning to find another downlink

channel. The SS shall determine from the channel description parameters whether it may use the uplink channel. If the channel is not suitable then the SS shall continue scanning to find another downlink channel. If the channel is suitable, the SS shall extract the parameters for this uplink from the UCD message. It then waits for the next DL-MAP message and extracts the time synchronisation from this message. Then, the SS shall wait for a bandwidth allocation map for the selected channel. It may begin transmitting uplink in accordance with the MAC operation and the bandwidth allocation mechanism.

16.9.4 Initial Ranging and Automatic Adjustments

Ranging is the process of acquiring the correct timing offset and power adjustments such that the SS's transmissions are aligned to a symbol that marks the beginning of a minislot boundary in SC and SCa PHY, or aligned with the BS receive frame for OFDM and OFDMA PHY, and received within the appropriate reception thresholds [3].

16.9.5 Ranging Parameter Adjustment

Adjustment of local parameters (e.g., transmit power) in an SS as a result of the receipt (or non-receipt) of RNG-RSP is considered to be implementation-dependent with the following restrictions:

1. All parameters shall be within the approved range at all times.
2. Power adjustment starts from the initial value selected with the algorithm unless a valid power setting is available from a non-volatile storage, in which case this value may be used as the starting point.
3. Power adjustment needs to be done by reducing or increasing the specified amount in response to RNG-RSP messages.
4. If, during initialization, power is increased to the maximum value (without a response from the BS) it shall wrap back to the minimum. On receiving RNG-RSP, the SS shall not transmit until the RF signal has been adjusted in accordance with the RNG-RSP and has stabilized.

16.9.6 Negotiate Basic Capabilities

The SS informs the BS of its basic capabilities by transmitting an SBC-REQ message immediately after completion of ranging, with its capabilities set to 'on'. The BS responds with an SBC-RSP message with the intersection of the SS's and the BS's capabilities set to 'on'.

16.9.7 SS Authorization and Key Exchange

The BS and SS shall perform authorization and key exchange.

16.9.8 Registration

Registration is the process by which the SS is permitted to attach into the network and a managed SS receives its secondary management CID and thus becomes manageable. To register with a BS, the SS shall send a REG-REQ message to the BS. The BS shall respond with a REG-RSP message.

16.9.9 IP Version Negotiation

The SS may include the IP version parameter in the REG-REQ to indicate which versions of IP it supports on the secondary management connection.

16.9.10 Establish IP Connectivity

At this point, the SS shall invoke DHCP mechanisms (Dynamic Host Configuration Protocol) in order to obtain an IP address and any other parameters needed to establish IP connectivity.

16.9.11 Establish Time of Day

The SS and BS need to have the current date and time. This is required for time-stamping logged events for retrieval by the management system.

16.9.12 Transfer Operational Parameters

After the successful DHCP configuration, the SS shall download the SS configuration File using Trivial File Transfer Protocol (TFTP) on the SS's secondary management connection if specified in the DHCP response. The 'siaddr' field of the DHCP response specifies the TFTP configuration file server. The SS shall use an adaptive timeout for TFTP based on binary exponential backoff. When the configuration file download has completed successfully, the SS shall notify the BS by transmitting a TFTP-CPLT message on the SS's primary management connection. Transmissions shall continue periodically until a TFTP-RSP message is received with 'OK' response from the BS or the SS terminates retransmission due to retry exhaustion [3].

16.9.13 Establish Provisioned Connections

After the transfer of operational parameters (for managed SS) or after registration (for unmanaged SS), the BS shall send DSA-REQ messages to the SS to set up connections for pre-provisioned service flows belonging to the SS. The SS responds with DSA-RSP messages.

In case of mesh mode, network entry and initialisation is different from the PMP mode.

16.10 RANGING

Ranging [3] [9] is a collection of processes by which the SS and BS maintain the quality of the RF communication link between them. Uplink ranging consists of two procedures—initial ranging and periodic ranging. Initial ranging allows an SS joining the network to acquire correct transmission parameters, such as time offset and transmission power level, so that the SS can communicate with the BS. Following initial ranging, periodic ranging allows the SS to adjust transmission parameters so that the SS can maintain uplink communications with the BS. The downlink burst profile is determined by the BS according to the quality of the signal that is received by each SS. To reduce the volume of uplink traffic, the SS monitors the carrier to interference and noise ratio (CINR) and compares the average value against the allowed range of operation. This region is bounded by threshold levels. If the received CINR goes outside the allowed operating region, the SS requests a change to a new burst profile using one of two methods. If the SS has been granted uplink bandwidth (a data grant allocation to the SS's basic CID), the SS shall send a DBPC-REQ message in that allocation. The BS responds with a DBPC-RSP message. If a grant is not available and the SS requires a more robust burst profile on the downlink, the SS shall send an RNG-REQ message in an initial ranging interval. With either method, the message is sent using the basic CID of the SS. The coordination of message transmits and receipt relative to actual change of modulation is different depending upon whether an SS is transitioning to a more or less robust burst profile. The SS applies an algorithm to determine its optimal burst profile in accordance with the threshold parameters established in the DCD message.

16.11 NETWORK ARCHITECTURE

The WiMAX forum [10] is developing an end-to-end network architecture, specifying the access to core systems and its functionalities. It contains procedures and protocols to support, e.g., mobility, security, internetworking and authentication to a WiMAX subscriber station. Figure 16.23 is the representation of the network reference model. The main entities for this model are Mobile Subscriber Stations (MSS), Access Service Network (ASN) and Connectivity Service Network (CSN). This reference model also contains

interfaces between the different entities. These interfaces define procedures and protocols and serve as logical, rather than physical, links across the entities.

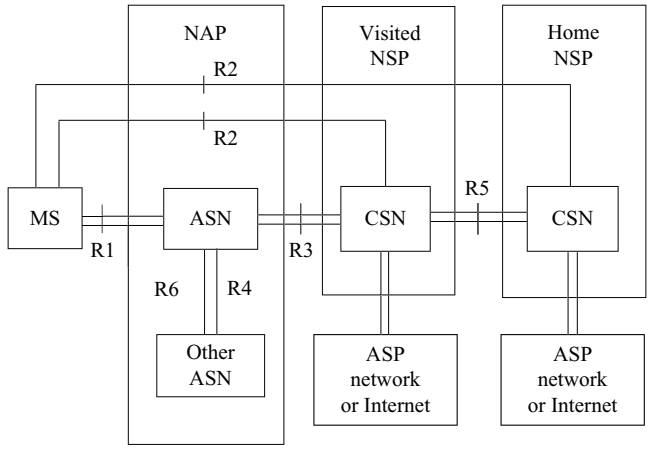


Fig. 16.23 *WiMAX network reference model*

16.11.1 Access Service Network: ASN

ASN is deployed by a business entity called Network Access Provider (NAP), which provides an SS/MSS with L2 connectivity to a WiMAX radio network and connects users to Network Service Providers (NSP) managing a CSN. The ASN gateway serves as the interconnection between ASN and CSN. ASN consists of one or more ASN Gateways and BSs. It handles radio coverage in a geographical area. The MAC access functionalities like paging, locating, Radio Resource Management (RRM) and mobility between base stations are managed by ASN, including support for mobile IP with foreign agent functionalities. The ASN is the management entity of the WiMAX radio links.

16.11.2 Connectivity Service Network: CSN

The CSN is deployed by a business entity called the **Network Service Provider (NSP)**. A CSN is a set of network functions that provide IP connectivity to WiMAX subscriber stations. The CSN contains gateways for Internet access, routers, servers or proxies for AAA, IP-allocation, user databases, and internetworking devices. It also handles admission and policy control, mobility between ASN and specific WiMAX services such as location-based services or law-enforcement services.

WiMAX subscribers may enter for contractual services—QoS, bandwidth, etc., with the NSP and access these services through the current ASN in which it is situated. The user can then use the service provider's network or roam to networks deployed by other companies as long as the home network has a roaming agreement with the visitor network. The foreign ASN uses either its own management functions of the foreign CSN, and proxies them to the home network, or communicates directly with the home network CSN.

Inside an ASN network entity there are at least one ASN gateway (ASN GW) and a base station (BS). The BS handles the connection to the MS while the ASN GW takes care of the contact with the CSN. An ASN GW can be associated with one or more BSs and a BS can have relations to one or more ASN GWs. This segmentation of the ASN enables multi-vendor systems where different vendors can produce different parts of the ASN and they still function together.

16.11.3 ASN Reference Points

To identify the different interfaces used to communicate within an ASN, the MS and the rest of the network, a number of reference points are introduced as shown in Fig. 16.24. These reference points define the set of protocols and procedures needed in the communication. Most of the reference points are logical mappings but when, as in the case of R1, the functional entities are in different physical devices, the reference point refers to a physical interface. R1 and R3 are the reference points used in communication with entities outside the ASN, while R6 and R8 are used inside an ASN. The R4 interface is used both inside and outside the ASN since it is the logical link between ASN GWs regardless of whether they are within the same ASN or in different ASNs. R1 is the physical interface between the MS and the serving BS and R3 is the logical link between ASN GW and CSN. The communication among BSs is handled through R8 while the BS-ASN GW interaction goes via R6.

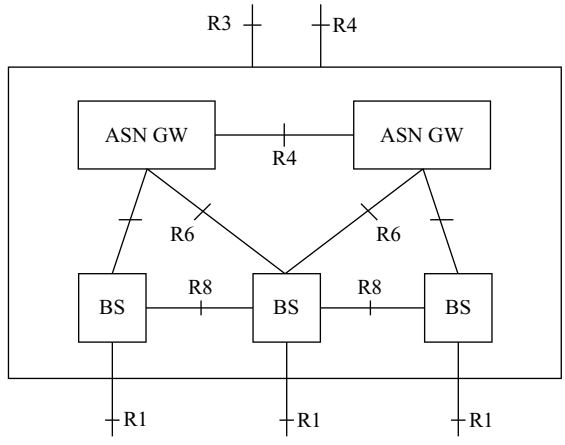


Fig. 16.24 *ASN reference points*

16.12 802.16e HANDOVER PROCEDURES

The 802.16e standard [4] can handle mobility in WiMAX. Like any other cellular networks, handover in WiMAX is performed with the help of layers 1, 2 and 3 together. It supports handovers through procedures and functions at the BS/MS level. The standard defines the means for gathering information and performing a handover but the decision whether to perform a handover or not is not defined. Although Layer 3 makes the ultimate decision of handover, the PHY and MAC layers play a major role by providing information about the handover and triggers for handover. For example, consider a situation when in order to perform a handover, the MS needs to switch to BS to receive higher signal quality or when the MS obtains improved QoS from another BS.

To perform a handover, the MS needs to acquire information about the network. This can be done with **network topology advertisements or scanning of neighbor BSs** with or without the optional association procedure. Network topology advertisements are broadcast messages sent out by all BSs. These messages contain information about neighboring BSs and their channels. This information will simplify the **MS synchronization with a new BS** since there is no need for the MS to listen to the target BS's DCD/UCD1 messages. The serving BS may receive the required information about its neighbors through the backbone.

Again the MS may use its time to scan its environment for potential target BSs to localize then and it may investigate the quality of their channels. The MS requests scanning intervals from the serving BS and may start scanning after getting permission from the serving BS. During the scanning interval, the serving BS assumes the MS to be in the scanning mode and it may buffer data incoming to the MS. When the MS leaves the scanning mode, the serving BS starts to send the buffered data to the MS. The MS can at any time terminate the scanning and may start to send PDUs again. When a BS receives PDUs from an MS supposed to be in the scanning mode, it will assume that the MS is no longer in the scanning mode and resume normal operation. To reduce overhead due to many scanning requests, the MS can ask for a group of scanning intervals. The intervals will be interleaved with periods of normal operation.

The association procedure is an optional feature in the standard and it can occur during scanning. There are three levels of association—scan/association without coordination, association with coordination and network-assisted association reporting. The goal of association is to enable the MS to collect and store

information about BSs. The gathered information is saved during a reasonable period of time and it can help the MS further in decisions regarding handovers.

16.12.1 Handover Process

The handover process consists of six different steps as indicated below.

1. Cell reselection
2. Handover decision and initiation
3. Synchronization to target BS downlink
4. Ranging
5. Termination of service
6. Handover cancelation

1. **Cell Reselection** It is the step where the MS acquires information about neighboring BSs in the network. The information is used in evaluation of the possibility to perform a handover. It can be done by using the information in the network topology advertisements or it may require a scanning interval to obtain the needed information. The cell reselection phase does not need to occur in relation to a handover decision.
2. **Handover Decision and Initiation** This is the decision to migrate the MS from the serving BS to a target BS. This decision can be triggered in the MS as well as in the BS. To commence the actual handover, the requesting party sends a handover request message that will trigger a sequence of handover-related messages to be sent between the MS and BS. When the MS takes a decision, it sends an MOB_MHSO-REQ message to the BS indicating one or more BSs as the targets for handover. The BS in reply sends an MOB_BSHO-RSP to the MS indicating the target BS to be used in the handover process. The MS then sends the MOB_MSHO-IND indicating which of the BSs indicated in the MOB_BSHO-RSP will be used for handover. On the other hand, when the BS takes the decision, it sends the MOB_BSHO-REQ message to MS indicating one or more BSs as the target. The MS in response sends the MOB_MSHO-IND message indicating the receipt of the target BS.
3. **Synchronization** To establish communication with the target BS, the MS needs to **synchronize** to its downlink channel. During this phase, the MS receives downlink and uplink transmission parameters. If the MS has previously received information about this BS (through the network topology acquisition) the length of this process can be shortened. The MS begins by processing the downlink frame preamble of the target BS. The DL frame preamble provides the MS with time and frequency synchronization with the target BS. When the MS is synchronized to the channel, it needs to perform initial ranging or handover ranging. MS decodes DL-MAP, UL-MAP, DCD and UCD messages to get information about the ranging channel.
4. **Ranging** It is a procedure where the MS receives the correct transmission parameters, e.g., time offset and power level. The MS uses the ranging channel to perform the initial ranging for synchronizing its UL transmission with the BS. The target BS may obtain information about the MS through the backbone, and depending on the target BSs knowledge about the MS, some parts of the ranging process may be omitted.
5. **Termination** It is the step after establishing the connection with the serving BS. The MS may decide to terminate the connection by sending an MOB_HO_IND message to the serving BS. The serving BS will terminate all connections associated with the MS and remove all information in queues, counters, etc.
6. **Cancellation** During the handover, the MS has the right to cancel the handover and resume normal communication with the serving BS. The only condition is that the MS does not try to cancel after a specified time has elapsed.

16.12.2 Types of Handover

The WiMAX architecture extends the 802.16 standard and that also includes the mechanisms for handovers. While the 802.16 standard provides support for handover between base stations, WiMAX offer protocols for handover higher up in the network structure. The WiMAX architecture supports mechanisms such as intra/inter ASN handover, roaming between NSPs, seamless handover at vehicular speed and micro/macro mobility. This section will study the architecture and its handover procedures more thoroughly with a focus on intra/inter ASN handovers. Depending on which role of a BS or ASN GW take on in a handover, they get different names.

Intra-ASN Handover The intra ASN handover is performed between BSs or within sectors in one BS belonging to the same ASN. The BSs can be connected to the same ASN GW or different ASN GWs within the same ASN. It will still be an intra-ASN handover. **If there is only one BS within an ASN, an intra-ASN handover cannot be performed** unless the BS has several antenna sectors. Intra-ASN handover is performed to minimise the delay and data loss during the MS's transition between BSs. If the MS uses services like IP or MIP there will be no need for a change of IP-address after the handover since the movement of the MS is not visible from outside the ASN. With respect to Fig. 16.24, the reference point involved for intra-ASN handovers are R6 and R8 and sometimes R4 when the target BS is connected to another ASN GW than the serving BS. Fig. 16.25 shows the intra-ASN handover scenario.

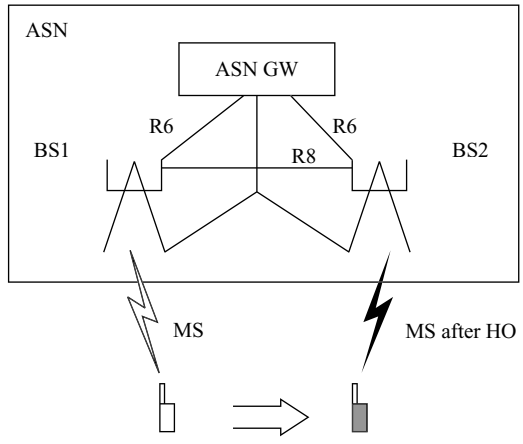


Fig. 16.25 *Intra-ASN handover*

Inter-ASN Handover An inter-ASN handover occurs between BSs when they are on different ASNs, as shown in Fig. 16.26. During an inter-ASN handover, the ASN GWs in separate ASNs need to coordinate their actions for smooth handover to the MS. There are two possible ways of dealing with the data during an inter-ASN handover—anchoring and re-anchoring. The purpose of anchoring is to avoid a path update and hence a redirection of the data path, where in the re-anchoring case update will be performed.

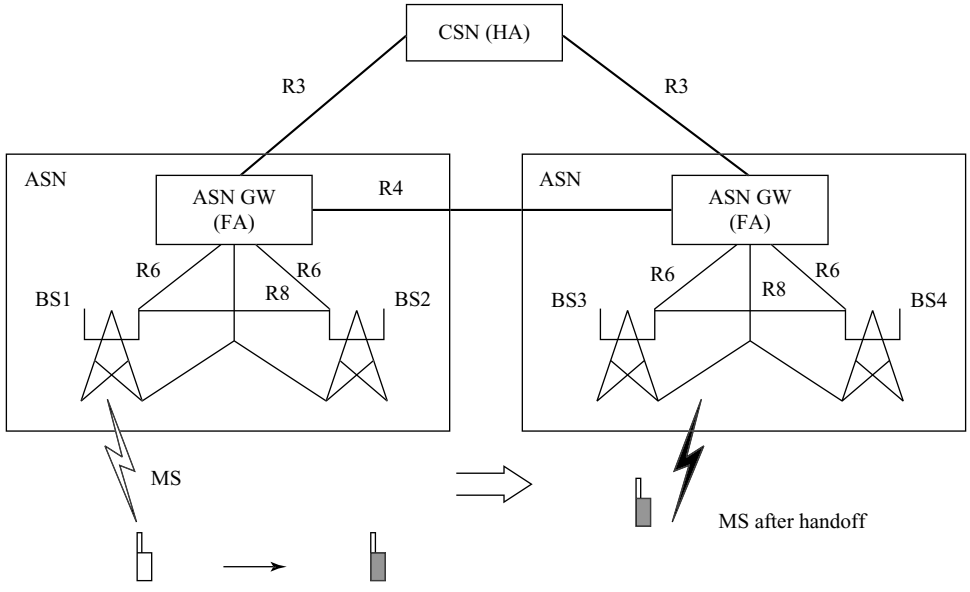


Fig. 16.26 *Inter-ASN handover scenario*

The decision to anchor or re-anchor the data path is made by the target or anchor ASN GW and there are three different decision procedures with two possible outcomes. Either both parties can decide that a re-anchoring is not needed or one of the ASN GW decides that it wants a re-anchoring. It is always the target ASN GW that will first make the decision. This decision is based upon its implementation dependence and is not included in the scope of the WiMAX document. Figure 16.26 shows the inter-ASN handover scenario.

Since in intra-ASN handover, no update for IP change is required during the handover, it falls under the category of handover without CoA update; whereas, in the inter-ASN handover, anchoring is used as a handover without CoA update but otherwise it is a handover with CoA update. ASN-anchored mobility is known as **intra-ASN** or **micromobility** where the data path uses the same ASN foreign agent. ASN anchoring involves migration of R6 reference point where R8 is used to transfer undelivered data after handoff.

CSN-anchored mobility is also called inter-ASN or macromobility. The MS moves from one ASN FA to a new ASN FA. The handover process occurs across R3 reference point with the tunneling over R4 to transfer undelivered data packets.

Summary

In this chapter, an overview of WiMAX broadband access technology has been provided. Starting from the spectrum allocation, physical layer and MAC layer functionalities have been discussed with all-important technical fundamentals. The network architecture, as provided by the WiMAX forum, is also described along with handover and mobility management. WiMAX has a very flexible MAC layer that can accommodate a variety of traffic types, including voice, video and multimedia and provide strong QoS. The WiMAX architecture supports Layer 2 and Layer 3 mobility. The detail of WiMAX is obtained in the standards given as references. Readers are advised to read those standards for in-depth knowledge of WiMAX.

References

[1] Scheibe, K.P., L.W. Carstensen Jr., T.R. Rakes, L.P. Rees, *Going the Last Mile: A Spatial Decision Support System for Wireless Broadband Communications*, Decision Support Systems 42 (2006) pp. 557—570, Elsevier.

[2] Chu, G., D. Wang and S. Mei, *A QoS Architecture for the MAC Protocol of IEEE 802.16 BWA System* Communications, Circuits and Systems and WestSino Expositions, IEEE 2002.

[3] *IEEE Standard 802.16^{TM}-2004. Part 16:Air Interface for Fixed Broadband Wireless Access Systems*, October 2004.

[4] *IEEE Standard 802.16e-2005. Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access System–Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Band*, December 2005.

[5] Ma, L., D. Jia, *The Competition and Cooperation of WiMAX*, *WLAN and 3G*', White paper, Beijing Consulting and Design Institute of P&T, P.R. China.

[6] Ghosh, A., R. David, J.G. Andrews and R. Chen, *Broadband Wireless Access with WiMAX/802.16: Current Performance Benchmarks and Future Potential*, IEEE Communication Magazine, pp. 129–136, Feb. 2005.

[7] Tsai, T.C., C.H. Jiang and C.Y. Wang, *CAC and Packet Scheduling Using Token Bucket for IEEE 802.16 Networks,* Journal of Communications, Vol 1, pp. 30–37, May 2006.

[8] Wang, S., K. Stanwood, Y. Bourlas, and R. Johnson, *ATM Convergence Sublayer for 802.16.1*, IEEE, 2004.

[9] Eklund, C., *IEEE Standard 802.16: A Technical Overview of the WirelessMAN™ Air Interface for Broadband Wireless Access*', IEEE Communication Magazine, June 2002.

[10] *WiMAX Forum: Recommendation and Requirements for Networks Based on WiMAX Forum certified™ products. Release 1.0*, February 23, 2006.
[11] Xiao, X., W. K.G. Seaw, Y.H. Chew and C.C. Ko, *Upstream Resource Reservation and Scheduling Strategies for Hybrid Fiber/Coaxial Networks*, CWC, Singapore.
[12] Molisch, A.F., *Wireless Communications,* Wiley Student Edition, 2005.

Questions for Self-Test

16.1 WiMAX is the new revolution in the area of wireless broadband services.
a. True b. False

16.2 The combining effect of broadband services in the wireless domain is the result of
a. fixed WiMAX b. mobile WiMAX

16.3 The advantages of WiMAX:
a. Scalable b. Low upgrade cost
c. Wider coverage d. Rapid deployment
e. All of the above

16.4 The first alternative for wireless access deployed for voice communication is
a. WiFi b. WLL

16.5 The new revolution in the area of wireless broadband services is
a. WiFi b. WiMAX

16.6 A first-generation broadband system at 2.5 GHz and 3.5 GHz is
a. LOS b. NLOS

16.7 The second-generation broadband systems use high-end digital signal processing to encounter multipath effect:
a. True b. False

16.8 IEEE 802.16a operates at 2-11GHz band for
a. NLOS
b. uses OFDM for PHY
c. all the above

16.9 IEEE 802.16e supports mobility, and uses scalable PHY.
a. True b. False

16.10 The most likely WiMAX frequency bands are
a. 2.3 GHz b. 2.5 GHz
c. 3.5 GHz d. 5.7 GHz

16.11 Indicate which one of these are licensed and unlicensed bands.
a. 3.5 GHz
b. 2.5–2.7 GHz
c. 5 GHz

16.12 The uneconomical broadband service for last mile solution is
a. fiber line b. WiMAX

16.13 WiMAX may be the alternative of fiber MAN in remote places.
a. True b. False

16.14 An important use of WiMAX is the cooperation with WiFi as a backhaul service.
a. True b. False

16.15 Throughput capacities of WiMAX are not dependent on channel bandwidth.
a. True b. False

16.16 Use of WiFi is simple for local area network.
a. Yes b. No

16.17 The advantage of WiMAX compared to 3G is the flexible selectable channel bandwidth.
a. True b. False

16.18 3G is more suitable than WiMAX in respect of mobility support.
a. True b. False

16.19 WiMAX downlink transmission is
a. unicast b. multicast

16.20 Uplink is always
a. unicast b. multicast

16.21 The word WiMAX stands for _____.

16.22 All communication in the WiMAX network is controlled by _____.

16.23 There is no _____ between the subscriber stations.

16.24 On the downlink data to subscriber, stations are multiples in _____.

16.25 The downlink in WiMAX is generally _____.

16.26 The connection in the downlink direction is either _____ or _____.

16.27 Uplink connections are always _____.

16.28 _____is used among multiple subscriber stations.

16.29 OFDM transmission is NLOS (Non-Line of Sight Propagation) because of _____ radio propagation environment.

16.30 In OFDM technology, the high data-rate stream is divided into several _____ data streams.

16.31 To avoid inter-symbol interference completely, OFDM uses guard interval between the symbols _____ than the expected multipath delay spread.

16.32 Implementation of OFDM is easy to use _____.

16.33 In mobile WiMAX, multiple access scheme known as _____ is used among multiple users.

16.34 Fixed WiMAX OFDM PHY uses _____ FFT, among them _____ subcarriers used for carrying data and the rest are used as pilot subcarriers and guard bands.

16.35 In Mobile WiMAX OFDMA PHY FFT size is variable from _____ to _____.

16.36 The subchannelisation scheme based on contiguous subcarriers in WiMAX is called _____.

16.37 Adaptive modulation and coding increases the overall system _____.

16.38 In PMP architecture, each SS has _____ bits unique MAC address to be used as equipment identifier.

16.39 An SS requests uplink bandwidth _____ basis, whereas the BS grants bandwidth to an SS as an _____.

16.40 In PMP mode, traffic only occurs between the _____ while in the mesh mode traffic can be routed through other SSs.

16.41 What are the different frequency bands available for WiMAX standard?

16.42 Compare and contrast WiFi-WiMAX-3G optical fiber deployment scenarios.

16.43 Compare WiMAX over other wideband frequency spectrum.

16.44 What are the three different terms of mesh systems? Give their definitions.

16.45 Why and how is guard band used in OFDM?

16.46 What is OFDMA? How is the guard band related with multipah delay spread?

16.47 WiFi and WiMAX may be the alternative solution for fixed broadband services in rural areas. Comment on this issue.

16.48 How is subchannelization performed in WiMAX PHY?

16.49 Describe the PHY layers for fixed and mobile WiMAX.
16.50 Discuss the multiple access technology used for WiMAX.
16.51 What is adaptive modulation and coding? How is it used in the WiMAX system?
16.52 What is the necessity of using bandwidth adaptive modulation and coding?
16.53 Discuss the PHY layer frame structure and basic access methods for WiMAX.
16.54 Outline the downlink and uplink frame structures in WiMAX standard.
16.55 What are the different sublayers of MAC layers in WIMAX? Discuss the functionalities of each layer.
16.56 Discuss the MAC PDU frame format in WiMAX. What are the different MAC subheaders available?
16.57 How is packing done for fixed-length and variable-length MAC SDUs into a single MAC PDU?
16.58 What are the different scheduling services defined in fixed and mobile WiMAX? Discuss in brief each of those service types.
16.59 With the help of a schematic diagram, discuss the WiMAX network reference models provided by the WiMAX forum.
16.60 What are the different steps of handover procedure in mobile WiMAX? Discuss the Inter-ASN handoff procedure in mobile WiMAX.

Brief Overview of 3G LTE

Appendix-A

The successful evolution and deployment of GSM family of technologies, generally known as 3GPP family have gone through the development phase of GSM, EDGE, UMTS, HSPA, HAPA^{+}, LTE and LTE-A. In the commercial market, HSPA^{+} continues its progress and in the way LTE revolution begun.

Long Term Evolution (LTE) is the next step forward in cellular 3G services. In the world of telecommunications, people today are more mobile and connecting themselves to mobile interfaces than ever. We have more new sophisticated handheld mobile devices to stay in touch with one another over wireless networks. The high speed mobility and on demand access of network and multimedia applications is the driving force for LTE technology development. LTE is an important technology transfer starting from circuit switch to packet switch and landed over to All-IP network architecture.

At the beginning of mobile broadband, AT&T launched UMTS enhanced with High Speed Downlink Packet Access (HSDPA) in 16 major markets throughout the U.S. in December 2005, and has become the first operator in the world to launch HSDPA on a wide-scale basis. AT&T deployed HSPA – capable of peak theoretical downloads speeds of up to 3.6 Mbps – in more than 350 U.S. cities and then upgraded its entire HSPA network to peak theoretical capabilities of up to 7.2 Mbps. AT&T announced plans to deploy HSPA+ in 2010, and began trials of LTE in the 700 MHz band with commercial deployment of LTE in and around 2011. The useful application of LTE is Voice Over LTE (VoLTE) expected to be available in the year 2013 [A.1].

LTE supports both frequency-division duplex (FDD) and time-division duplex (TDD). In order to support large number of different spectrum allocations, LTE supports also a wide range of system bandwidths. LTE aims for a smooth evolution from earlier 3GPP systems such as time division synchronous code division multiple access (TDSCDMA) and wide-band code division multiple access/high-speed packet access (WCDMA/HSPA), as well as 3GPP2 systems such as code division multiple access (cdma) 2000 [A.2].

The LTE as defined by the 3rd Generation Partnership Project (3GPP) is a highly flexible radio interface, which is using the VoIP to transmit the voice services and packet the data for all services. The first release of LTE was published in March 2009 and is referred to as LTE Release-8. LTE Release- 8 provides high peak data rates of 300 Mb/s on the downlink and 75 Mb/s on the uplink for a 20 MHz bandwidth. In LTE Release 8, orthogonal frequency-division multiplexing (OFDM) is the DL (Downlink) multiple access scheme, while single-carrier frequency-division multiple access (SC-FDMA) is the UL (uplink) multiple access scheme. LTE Release-8 also supports scalable bandwidth up to 20 MHz, and uses DL/UL frequency selective and DL frequency diverse scheduling, respectively. The DL subframe structure is common to both TDD and FDD. The Media Access Control Layer (MAC layer) in data link layer of Open System Interconnection (OSI) aims to control the authority of user about the accessing media and resource.

LTE has many technical benefits to cellular networks. Bandwidth is scalable from 1.25 MHz to 20 MHz. This will suit the needs of different network operators that have different bandwidth allocations, and also allow operators to provide different services based on spectrum [A.3]. LTE is also expected to improve spectral efficiency in 3G networks, allowing carriers to provide more data and voice services over a given bandwidth. LTE encompasses high-speed data, multimedia unicast and multimedia broadcast services. The LTE PHY is a highly efficient means of conveying both data and control information between an enhanced base station (eNodeB) and mobile user equipment (UE).

The LTE PHY is designed to meet the following goals [A.4]:

1. **Support scalable bandwidths of 1.25, 2.5, 5.0, 10.0 and 20.0 MHz**
2. **Peak data rate that scales with system bandwidth**
   (a) Downlink (2 Ch MIMO) peak rate of 100 Mbps in 20 MHz channel
   (b) Uplink (single Ch Tx) peak rate of 50 Mbps in 20 MHz channel
3. **Supported antenna configurations**
   (a) Downlink: 4×2, 2×2, 1×2, 1×1
   (b) Uplink: 1×2, 1×1
4. **Spectrum efficiency**
   (a) Downlink: 3 to $4 \times$ HSDPA Rel. 6
   (b) Uplink: 2 to $3 \times$ HSUPA Rel. 6
5. **Latency**
   (a) C-plane: <50 – 100 msec to establish U-plane
   (b) U-plane: <10 msec from UE to server
6. **Mobility**
   (a) Optimized for low speeds (<15 km/hr)
   (b) High performance at speeds up to 120 km/hr
   (c) Maintain link at speeds up to 350 km/hr
7. **Coverage**
   (a) Full performance up to 5 km
   (b) Slight degradation 5 km – 30 km
   (c) Operation up to 100 km should not be precluded by standard

LTE Architecture

3GPP has developed the 4G cellular networks based on LTE standard, which support the OFDM technology for downlink transmission and Single carrier FDM for uplink transmission for supporting the broadband data communications. MIMO (Multiple Input Multiple Output) Technique is used in the DL of LTE Rel 8. Figure A.1 denotes the basic architecture of LTE [A.5]. The base stations for LTE is denoted by eNode-B (eNB) and the mobile stations or terminals as UE. There are pico cells, femtocells and home eNBs along with relay nodes. The Relay Node (RN) is served by eNBs called donor eNB (DeNB). The same eNB may be served as the DeNB or the regular eNB. MME is the mobility management entity for control plane and S-GW is called the local mobility anchor points for the data plane. The X2 interface is defined as a direct eNB to eNB interface allows for inter cell interference coordination (ICIC) and s_1 is the secondary synchronization signal.

Picocells are regular eNBs with the only difference of having lower transmitted power than traditional macro cells. They are, typically, equipped with omni-directional antennas (not sectorized) and are deployed indoors or outdoors often in a planned (hot-spot) manner. Their transmit power ranges from 250 mW to approximately 2 W for outdoor deployments, while it is typically 100 mW or less for indoor deployments [A.6].

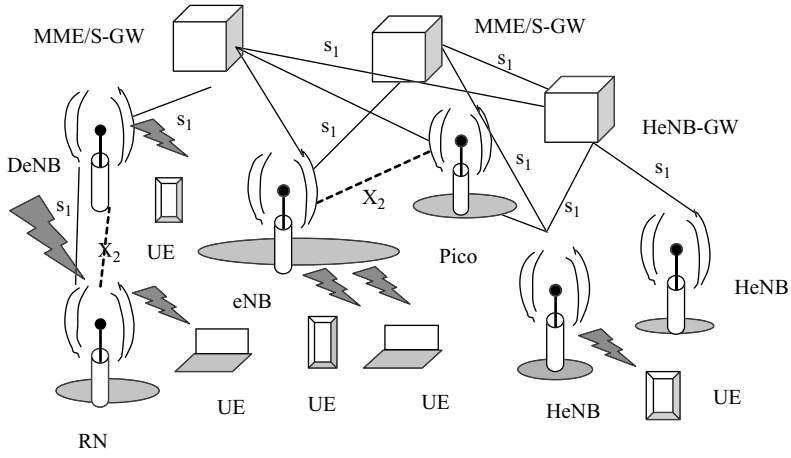


Fig A.1 *Basic components for 3G LTE Architecture*

Femtocells are generally consumer deployed (unplanned) network nodes for indoor application. It is used with a network backhaul facilitated by the consumer's home DSL (digital subscriber line) or cable modem. Femtocells are typically equipped with omnidirectional antennas, and their transmit power is 100 mW or less [A.6]. Femto cells are classified as open or closed. Closed femtos restrict the access to a closed subscriber group (CSG), while open femtos are similar to picocells but with the network backhaul provided by the home DSL or cable modem.

A **relay node** (RN) is a network node without a wired backhaul. The backhaul, which provides the attachment of the RN to the rest of the network, is wireless and uses the air interface resources of the wireless system. Relays typically yield performance improvement by providing higher throughputs to UE that would otherwise be located in poor geometry locations with respect to the macrocell sites. There are two types of relays being discussed in the context of 3GPP standards, Type 1 and Type 2 relays. Type 1 and Type 2 relays have the following characteristics:

Type 1

1. A relay cell should have its own physical cell ID and transmit its own synchronization channels, reference symbols, and so on, and will be distinct from the donor cell.
2. The UE should receive scheduling information and HARQ feedback directly from the RN and send its control channels (SR/CQI/ACK) to the RN.
3. The RN appears as a Release 8 eNB to Release 8 UE.

Type 2

1. The RN does not have a separate Physical Cell ID and thus would not create any new cells.
2. It is transparent to Release 8 UE; Release 8 UE is not aware of its presence.
3. It can transmit physical downlink shared channel (PDSCH) but does not transmit CRS and PDCCH.

The basic time and frequency unit in the DL (UL) is one OFDM (SC-FDM) symbol and one subcarrier (virtual subcarrier), respectively. The subcarrier spacing is 15 kHz and therefore, the OFDM symbol duration is 66.67 μs. Each OFDM/SC-FDM symbol is pre-appended with a cyclic prefix (CP) to suppress the inter-symbol interference and mitigate multi-path. Two CP durations are defined; the normal CP has duration of 4.7 μs and the extended CP has duration of 16 μs. One resource element corresponds to one sub-carrier (virtual sub-carrier) in one OFDM (SC-FDM) symbol. OFDM (SCFDM) symbols are grouped in subframes of 1 ms duration. Each subframe is composed of two 0.5 ms slots [A.3] as shown in Fig. A.2.

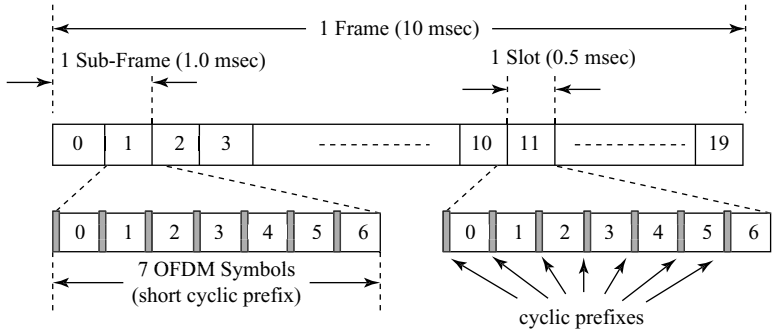


Fig A.2 *OFDM symbol for 3G LTE*

LTE frames are 10 msec in duration. They are divided into 10 subframes, each subframe being 1.0 msec long. Each subframe is further divided into two slots, each of 0.5 msec duration. Slots consist of either 6 or 7 ODFM symbols, depending on whether the normal or extended cyclic prefix is employed [A.3].

LTE Advanced (LTE-A)

LTE-Advanced (also known as LTE Release 10) significantly enhances the existing LTE Release 8 and supports much higher peak rates, higher throughput and coverage, and lower latencies, resulting in a better user experience. These requirements are met using a variety of techniques, including [A.9]:

1. Carrier aggregation
2. DL spatial multiplexing using up to eight-layer multiple-input multiple-output (MIMO)
3. DL intracell CoMP transmission and reception
4. UL Spatial Multiplexing using four-layer MIMO

In addition to the improvement in spectral efficiency, substantial reduction in latency is also targeted. The goals are to reduce the transition time from idle to connected mode from 100 ms in LTE to less than 50 ms in LTE-A. Similarly, the transition from dormant to active should be reduced from 50 ms in LTE to less than 10 ms in LTE-A. In LTE-A capacity and coverage enhancement can also be achieved using a heterogeneous network, which is a collection of low-power nodes distributed across a macrocell (homogeneous) network. There are various types of low-power nodes including microcells, picocells, femtocells, and relays. These low-power nodes are deployed in various environments including hot spots, homes, enterprise environments, and low geometry locations to improve the overall capacity and coverage of the system.

Power Control in 3G LTE

Power control refers to setting output power levels of transmitters, that is, base stations in downlink and mobile stations in uplink, with an objective to improve system capacity, coverage and user quality (data rate or voice quality), and to reduce power consumption. To reach these objectives, power-control mechanisms typically aim at maximizing the received power of desired signals, while limiting the generated interference. In the downlink, a simple and efficient power control strategy, used in most recent system concepts,

is to transmit with a constant output power. Often the maximum base station power is used. Variations in channel conditions and interference levels are adapted too by means of scheduling and link adaptation rather than with power control. This strategy obviously maximizes the received power. The generated interference power is instantaneously high, but maximizing the data rate, thus minimizing the transmission time, through scheduling and link adaptation, minimizes the interference energy generated for a given amount of data transferred. The LTE uplink power control may be considered as a means to apply this downlink concept in the uplink direction.

Carrier Aggregation

Bandwidth extension in LTE-A is supported via carrier aggregation. Carrier aggregation allows deployment bandwidths of up to 100 MHz, enabling peak target data rates in excess of 1 Gb/s in the DL and 500 Mb/s in the UL to be achieved. In addition, it can be used to effectively support different component carrier types that may be deployed in heterogeneous networks. Carrier aggregation is attractive because it allows operators to deploy a system with extended bandwidth by aggregating several smaller component carriers while providing backward compatibility to legacy users. Three different types of component carriers are envisioned [A.9]:

1. **Backward-compatible carrier:** All LTE UE can access this type of carrier regardless of the supported release. In this case all of the current LTE features must be supported.
2. **Non-backward-compatible carrier:** Only LTE-A UEs can access this type of carrier. This carrier may support advanced features such as control-less operations or the anchor-carrier concept not available to LTE UE.
3. **Extension carrier:** This type of carrier operates as an extension of another carrier. As an example, to provide services to home eNBs in the presence of high interference from the macrocell. UE can only access this type of carrier as part of a carrier aggregation set.

Conclusion LTE is a newly emerged technology therefore every fields of this technology needs a detail study and extensive research work. In today's world, a technology is more successful which consumes less energy but delivers high quality of services. In LTE, using various algorithms, power consumption can be controlled at the same time increasing its coverage and capacity. Very brief overview of LTE technology is discussed. For more details, some important references are provided.

References:

[A.1] Report, 4G Mobile Broadband Evolution: 3GPP Release 10 and Beyond, www.4G Americas.org, Feb 2011.

[A.2] Astély, David, Erik Dahlman, Anders Furuskär, Ylva Jading, Magnus Lindström, and Stefan Parkvall, Ericsson Research, "LTE: The Evolution of Mobile Broadband", *IEEE Communication Magazine, pp 41-51, April 2009*.

[A.3] Jim Zyren, Wes McCoy, "Overview of the 3GPP Long Term Evolution Physical Layer", *White Paper, Freescale Semiconductor, July 2007*.

[A.4] 3GPP TR 25.913 - v7.3.0, *Requirements for EUTRA and EUTRAN*, http://www.3gpp.org/ftp/Specs/archive/25%5Fseries/25.913/

[A.5] 3GPP Evolved Universal Terrestrial Radio Access (EUTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall Description; Stage 2 (Release 10).

[A.6] Damnjanovic, Aleksandar, An Montojo, Yongbin Wei, Tingfang Ji, Tao Luo, Madhavan Vajapeyam, Taesang Yoo, Osok Song, and Durga Malladi, "A Survey on 3GPP Heterogeneous Networks", IEEE Wireless Communication Magazine, pp. 9 –21, June 2011.

[A.7] A. Bou Saleh *et al*., "Comparison of Relay and Pico eNB Deployments in LTE-Advanced," *IEEE VTC 2009*.

[A.8] 3GPP TS 36.300 v8.0.0, *E-UTRA and E-UTRAN Overall Description; Stage 2,* http://www.3gpp.org/ftp/Specs/archive/36%5Fseries/36.300/

[A.9] Amitava Ghosh, Rapeepat Ratasuk, Bishwarup Mondal, Nitin Mangalvedhe, And Tim Thomas, "LTE-Advanced: Next-Generation Wireless Broadband Technology", *IEEE Wireless Communication, June 2010*.

Overview of Bluetooth Technology

Appendix-B

Bluetooth is the short-range radio link technology developed with the intention to replace the cable connecting portable and/or fixed electronic devices. Bluetooth allows for the replacement of the many propriety cables that connect one device to another with one **universal radio link**. Bluetooth radio modules operate in the unlicensed ISM band at 2.4GHz, and avoid interference from other signals by hopping to a new frequency after transmitting or receiving a packet. The beauty of this technology is the robustness, low complexity, low power and low cost and designed to operate in noisy frequency environments.

The research on Bluetooth was initiated at Ericsson of Sweden in 1994. The idea of Bluetooth comes from the desire to connect cellular phones with other devices without a cable. The name "Bluetooth" came after the 10th century Danish king of Denmark Harald Bluetooth [B.1].

Bluetooth technology is a short-range communications technology that is simple, secure, and everywhere ranging from mobile phones and computers to medical devices and home entertainment products. It is also low cost and consumes low power. Two Bluetooth enabled devices connecting to each other formed a pair. Any Bluetooth enabled device, almost everywhere in the world, can connect to other Bluetooth enabled devices located in proximity to one another.

Connections between Bluetooth enabled electronic devices allow these devices to communicate wirelessly through short-range, ad hoc networks known as piconets. A Piconet session is a communication link that must be created between devices to communicate with each other. Piconets are established dynamically and automatically as Bluetooth enabled devices enter and leave radio proximity. Each device in a piconet can also simultaneously communicate with up to seven other devices within that single piconet and each device can also belong to several piconets simultaneously [B.2].

A fundamental strength of Bluetooth wireless technology is the ability to simultaneously handle data and voice transmissions, which provides users with a variety of innovative solutions such as hands-free headsets for voice calls, printing and fax capabilities, and synchronization for PCs and mobile phones etc.

The range of Bluetooth technology is application specific. The Core Specification requires a minimum range of 10 meters/ 30 feet, but there is no set limit and manufacturers can tune their implementations to provide the range needed to support the use cases for their solutions.

If two devices come in contact with each other within 30 feet, the user will be prompted to initiate a communication session. Users then can either deny or accept the request to initiate a session. Only devices approved by the user can take part in the session. Data will appear as noise to unauthorized devices providing a great security feature.

The extension of Bluetooth is the new standard for Wireless Personal Area Network (WPAN)-IEEE802.15. The key limitations so far is its speed, the maximum data rate of 720kbps, so it cannot be used to connect DVD players or HDTV, and it takes a long time to transfer large picture files to a printer. Though new version of Bluetooth may address this issue and have much higher data rate.

Some of the key technical features of Bluetooth are given under.

Technical Challenges	Solution
Global operation	2.4GHz ISM band
Interference from other devices in ISM band and other Bluetooth Devices	Frequency Hopping Spread Spectrum (FHSS), Error correction coding
Low power consumption	Power control, Power-saving modes, Programmable packet length, Moderate data rate
Low Cost	FHSS, TDMA, Low receiver sensitivity
Security	FHSS, Link Layer Security using authentication and Encryption
Voice and Data Support	Using Circuit and Packet Switching
High error probability of wireless link	Controlled by using ARQ, FEC and CVSD

Bluetooth devices use a protocol called Frequency-Hopping Spread Spectrum. Sends data using packet switching on a range of frequencies. In each session one device is a master and the others are slaves. The master device decides at which frequency data will travel. Transceivers "hop" among 79 different frequencies in the 2.4 GHz band at a rate of 1600 frequency hops per second. The master device tells the slaves at what frequency data will be sent. This technique allows devices to communicate with each other more securely. The hopping sequence is calculated using the master's Bluetooth Device Address.

In the 2.4GHz ISM band, the use of spread spectrum is must to avoid interference. Although DSSS (Direct Sequence Spread Spectrum) can achieve higher data rate (11Mbps for 802.11b standard), FHSS has its advantage of low cost, low power, better security. FHSS also handles near-far problem better, since it will effectively block out-of-band signals. Considering the possible applications of Bluetooth, FHSS is a better solution. Figure [B.1] shows the FHSS communication between two bluetooth devices [B.5].

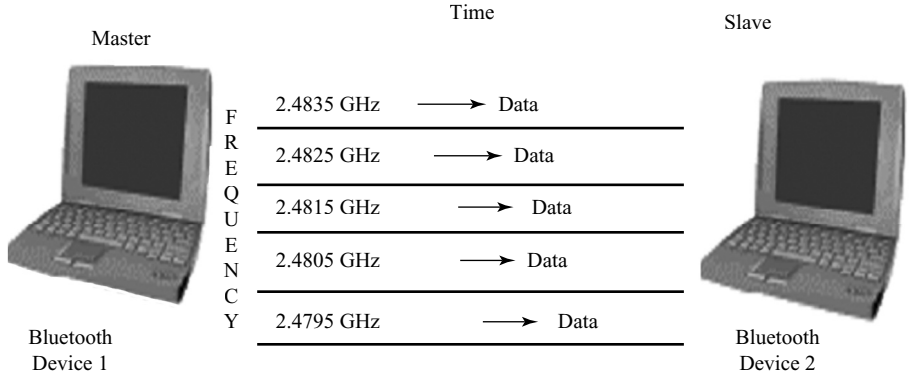


Fig B.1 *FHSS Communication in between two Bluetooth Devices*

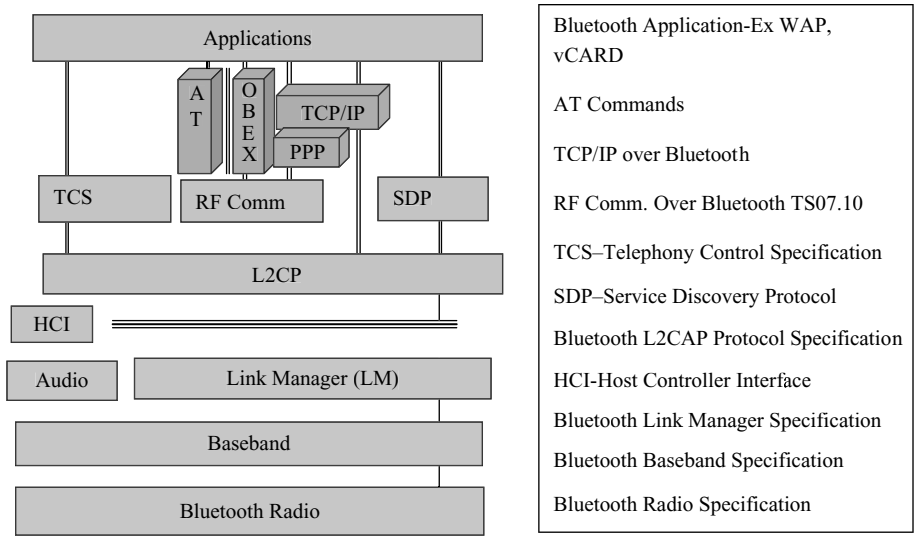


Fig B.2 *Bluetooth Architecture*

Figure B.2 Shows the Bluetooth protocol stack. The radio layer defines the technical characteristics of the Bluetooth radios. A Bluetooth radio operates on the 2.4 GHz ISM band which is license-free. It employs a 1,600 hops/sec frequency hoping, spread-spectrum (FHSS) technique. The radio hops in a pseudo-random fashion on 79, 1 MHz channels. The frequencies are located at (2,402+k) MHz, $k = 0, 1, \ldots, 78$. The modulation technique is a binary Gaussian frequency shift-keying (GFSK) and the baud rate is 1 Msymbols/sec.

The baseband defines the main procedures to enable devices to communicate with each other using the Bluetooth wireless technology. The baseband defines the Bluetooth piconets and their formation, and the Bluetooth links. A piconet is a collection of Bluetooth devices that can communicate with each other. A piconet contains at least one device identified as the master of the piconet and at most 7 other devices identified as slaves with which the master is actively involved in communications. A Bluetooth radio may serve either as a master or slave at different times and the master controls the communications.

With *time-division duplex* (TDD) master and slaves alternate transmit opportunities. Generally, the master transmits on even numbered slots defined by the master's Bluetooth clock, while the slaves transmit on odd numbered slots with each slot time 625 μ Sec.

The link manager protocol is a transactional protocol between two link management entities in communicating Bluetooth devices whose responsibility is to set-up the properties of the Bluetooth link. Through LMP transactions, a device may authenticate another one through a challenge response mechanism. Bluetooth devices may be authenticated and links may be encrypted. The authentication starts with the transmission of an LMP challenge packet.

HCI is an interface for host devices to access the lower layers of the Bluetooth stack through a standardized interface. Through the HCI, a host device passes and receives data destined to or coming from another Bluetooth device. The L2CAP layer shields the specifics of the Bluetooth lower layers and provides a packet interface to higher layers. At the L2CAP layer, the concepts of master and slave devices do not exist anymore. L2CAP traffic flows over logical channels terminating at the L2CAP layer of communicating devices. Channels may either be connectionless, or connection oriented. L2CP duties include protocol multiplexing, segmentation and reassembly of upper layer PDUs of length up to 64KB, QoS support and Groups abstraction.

Using SDP protocol a Bluetooth device can inquiry of the services that another device across a Bluetooth link may have and learn about how to get access to it. The SDP only provides information about services; it

does not provide access to them. SDP packets are carried over connection-oriented L2CAP channels between communicating devices.

The RFCOMM protocol is an important layer that is used to expose a serial interface to the packet based Bluetooth transport layers. In particular, the RFCOMM layer emulates the signals on the nine wires of an RS-232 interconnect cable.

Telephony control can be performed using the AT command set. Since, the AT commands have been designed to be passed over serial lines, Bluetooth devices use the RFCOMM to send and receive control signalling based on the AT command set. For example, using these commands, a dialer application in a notebook computer may instruct a cellular phone to dial-up an ISP location.

The Bluetooth specification comprises not only communications protocols but applications as well. Some of the potential markets include Digital mobile phones, Digital cordless phones, Wireless headsets, Data access points (hot spots), Laptop, desktop and PDA, Computer peripherals, Digital cameras and Home networking. Use of Bluetooth may appear as "hot-spots" in some hotels, shopping malls, airports. Compared to other similar wireless technologies, the biggest advantage is the low power and low cost, which makes it suitable for mobile applications. But because of its low data rate, high-speed applications like real time video would be a problem. A higher data rate version of Bluetooth is under discussion and development [B.3].

References:

[B.1] Bisdikian, Chatschik, "An Overview of the Bluetooth Wireless Technology", *IBM Research Division, RC 22109* (W0107-009) 6 June 2001.

[B.2] Jaap C. Haartsen, BLUETOOTH—The Universal Radio Interface for Ad Hoc, *Wireless Connectivity, Ericsson Review No. 3, 1998*.

[B.3] Wang**,** Hongfeng, "Overview of Bluetooth Technology"**,** Dept. of Electrical Engineering State College, July 3, 2001.

[B.4] http://www.bluetooth.com.

[B.5] http://www.xircom.com

[B.6] Robert Morrow, *Bluetooth Operation and use*, McGrawHill, 2002.

The Erlang-B and Erlang-C Tables

Appendix C

Table 1 *Offered Traffic Load in Erlangs in an Erlang-B System (Number of Channels from 2 to 17)*

Channel Number	Call Blocking Probability							
	0.001	**0.002**	**0.005**	**0.010**	**0.020**	**0.050**	**0.070**	**0.100**
2	0.046	0.065	0.105	0.152	0.223	0.381	0.470	0.595
3	0.194	0.249	0.349	0.455	0.602	0.899	1.057	1.271
4	0.439	0.535	0.701	0.869	1.092	1.525	1.748	2.045
5	0.762	0.900	1.132	1.361	1.657	2.218	2.504	2.881
6	1.146	1.325	1.622	1.909	2.276	2.960	3.305	3.758
7	1.579	1.798	2.157	2.501	2.935	3.738	4.139	4.666
8	2.051	2.311	2.730	3.127	3.627	4.543	4.999	5.597
9	2.557	2.855	3.333	3.783	4.345	5.370	5.879	6.546
10	3.092	3.426	3.961	4.461	5.084	6.216	6.776	7.511
11	3.651	4.021	4.610	5.160	5.841	7.076	7.687	8.487
12	4.231	4.637	5.279	5.876	6.615	7.950	8.610	9.474
13	4.830	5.270	5.964	6.607	7.401	8.835	9.543	10.470
14	5.446	5.919	6.663	7.352	8.200	9.729	10.485	11.473
15	6.077	6.582	7.375	8.108	9.010	10.633	11.434	12.484
16	6.722	7.258	8.100	8.875	9.828	11.543	12.390	13.500
17	7.378	7.946	8.834	9.652	10.656	12.461	13.353	14.522

Table 2 *Offered Traffic Load in Erlangs in an Erlang-B System (Number of Channels from 18 to 52)*

Channel Number	Call Blocking Probability							
	0.001	0.002	0.005	0.010	0.020	0.050	0.070	0.100
18	8.046	8.644	9.578	10.437	11.491	13.385	14.321	15.548
19	8.724	9.351	10.331	11.230	12.333	14.315	15.294	16.579
20	9.411	10.068	11.092	12.031	13.182	15.249	16.271	17.613
21	10.108	10.793	11.860	12.838	14.036	16.189	17.253	18.651
22	10.812	11.525	12.635	13.651	14.896	17.132	18.238	19.692
23	11.524	12.265	13.416	14.470	15.761	18.080	19.227	20.737
24	12.243	13.011	14.204	15.295	16.631	19.031	20.219	21.784
25	12.969	13.763	14.997	16.124	17.505	19.985	21.214	22.833
26	13.701	14.522	15.795	16.959	18.383	20.943	22.212	23.885
27	14.439	15.285	16.598	17.797	19.265	21.904	23.213	24.939
28	15.182	16.054	17.406	18.640	20.150	22.867	24.216	25.995
29	15.930	16.828	18.218	19.487	21.039	23.833	25.221	27.053
30	16.684	17.606	19.034	20.337	21.932	24.802	26.228	28.113
31	17.442	18.389	19.854	21.191	22.827	25.773	27.238	29.174
32	18.205	19.175	20.678	22.048	23.725	26.746	28.249	30.237
33	18.972	19.966	21.505	22.909	24.626	27.721	29.262	31.301
34	19.742	20.761	22.336	23.772	25.529	28.698	30.277	32.367
35	20.517	21.559	23.169	24.638	26.435	29.677	31.293	33.434
36	21.296	22.361	24.006	25.507	27.343	30.657	32.311	34.503
37	22.078	23.166	24.846	26.378	28.253	31.640	33.330	35.572
38	22.864	23.974	25.689	27.252	29.166	32.623	34.351	36.643
39	23.652	24.785	26.534	28.129	30.081	33.609	35.373	37.715
40	24.444	25.599	27.382	29.007	30.997	34.596	36.396	38.787
41	25.239	26.416	28.232	29.888	31.916	35.584	37.421	39.861
42	26.037	27.235	29.085	30.771	32.836	36.574	38.446	40.936
43	26.837	28.057	29.940	31.656	33.758	37.565	39.473	42.011
44	27.641	28.881	30.797	32.543	34.682	38.557	40.501	43.088
45	28.447	29.708	31.656	33.432	35.607	39.550	41.529	44.165
46	29.255	30.538	32.517	34.322	36.534	40.545	42.559	45.243
47	30.066	31.369	33.381	35.215	37.462	41.540	43.590	46.322
48	30.879	32.203	34.246	36.108	38.392	42.537	44.621	47.401
49	31.694	33.039	35.113	37.004	39.323	43.534	45.653	48.481
50	32.512	33.876	35.982	37.901	40.255	44.533	46.687	49.562
51	33.331	34.716	36.852	38.800	41.189	45.532	47.721	50.643
52	34.153	35.558	37.724	39.700	42.124	46.533	48.755	51.726

Table 3 *Offered Traffic Load in Erlangs in an Erlang-B System (Number of Channels from 53 to 87)*

Channel Number	Call Blocking Probability							
	0.001	0.002	0.005	0.010	0.020	0.050	0.070	0.100
53	34.977	36.401	38.598	40.602	43.060	47.534	49.791	52.808
54	35.803	37.247	39.474	41.505	43.997	48.536	50.827	53.891
55	36.630	38.094	40.351	42.409	44.936	49.539	51.864	54.975
56	37.460	38.942	41.229	43.315	45.875	50.543	52.901	56.059
57	38.291	39.793	42.109	44.222	46.816	51.548	53.940	57.144
58	39.124	40.645	42.990	45.130	47.758	52.553	54.978	58.229
59	39.959	41.498	43.873	46.039	48.700	53.559	56.018	59.315
60	40.795	42.353	44.757	46.950	49.644	54.566	57.058	60.401
61	41.633	43.210	45.642	47.861	50.589	55.573	58.098	61.488
62	42.472	44.068	46.528	48.774	51.534	56.581	59.139	62.575
63	43.313	44.927	47.416	49.688	52.481	57.590	60.181	63.663
64	44.156	45.788	48.305	50.603	53.428	58.599	61.223	64.750
65	44.999	46.650	49.195	51.518	54.376	59.609	62.266	65.839
66	45.845	47.513	50.086	52.435	55.325	60.619	63.309	66.927
67	46.691	48.378	50.978	53.353	56.275	61.630	64.353	68.016
68	47.540	49.243	51.872	54.272	57.226	62.642	65.397	69.106
69	48.389	50.110	52.766	55.191	58.177	63.654	66.442	70.196
70	49.239	50.979	53.662	56.112	59.129	64.667	67.486	71.286
71	50.091	51.848	54.558	57.033	60.082	65.680	68.532	72.376
72	50.944	52.718	55.455	57.956	61.035	66.694	69.578	73.467
73	51.799	53.590	56.354	58.879	61.990	67.708	70.624	74.558
74	52.654	54.463	57.253	59.803	62.945	68.722	71.671	75.649
75	53.511	55.337	58.153	60.727	63.900	69.738	72.718	76.741
76	54.368	56.211	59.054	61.653	64.857	70.753	73.765	77.833
77	55.227	57.087	59.956	62.579	65.814	71.769	74.813	78.925
78	56.087	57.964	60.859	63.506	66.771	72.786	75.861	80.018
79	56.948	58.842	61.763	64.434	67.729	73.803	76.909	81.110
80	57.810	59.720	62.667	65.363	68.688	74.820	77.958	82.203
81	58.673	60.600	63.573	66.292	69.647	75.838	79.007	83.297
82	59.537	61.480	64.479	67.222	70.607	76.856	80.057	84.390
83	60.403	62.362	65.386	68.152	71.568	77.874	81.106	85.484
84	61.268	63.244	66.294	69.084	72.529	78.893	82.156	86.578
85	62.135	64.127	67.202	70.016	73.490	79.912	83.207	87.672
86	63.003	65.011	68.111	70.948	74.452	80.932	84.258	88.766
87	63.872	65.896	69.021	71.881	75.415	81.952	85.309	89.861

Table 4 *Offered Traffic Load in Erlangs in an Erlang-B System (Number of Channels from 88 to 200)*

Channel Number	Call Blocking Probability							
	0.001	0.002	0.005	0.010	0.020	0.050	0.070	0.100
88	64.742	66.782	69.932	72.815	76.378	82.972	86.360	90.956
89	65.612	67.668	70.843	73.749	77.342	83.993	87.411	92.051
90	66.484	68.556	71.755	74.684	78.306	85.014	88.463	93.146
91	67.356	69.444	72.668	75.620	79.270	86.035	89.515	94.242
92	68.229	70.333	73.581	76.556	80.236	87.057	90.568	95.338
93	69.103	71.222	74.495	77.493	81.201	88.079	91.620	96.434
94	69.978	72.113	75.410	78.430	82.167	89.101	92.673	97.530
95	70.853	73.004	76.325	79.367	83.133	90.123	93.726	98.626
96	71.729	73.895	77.241	80.306	84.100	91.146	94.779	99.722
97	72.606	74.788	78.157	81.245	85.068	92.169	95.833	100.819
98	73.484	75.681	79.074	82.184	86.035	93.193	96.887	101.916
99	74.363	76.575	79.992	83.124	87.003	94.216	97.941	103.013
100	75.242	77.469	80.910	84.064	87.972	95.240	98.995	104.110
105	79.649	81.951	85.509	88.773	92.821	100.364	104.269	109.598
110	84.072	86.448	90.121	93.493	97.678	105.494	109.549	115.089
115	88.511	90.960	94.746	98.223	102.544	110.630	114.832	120.584
120	92.964	95.484	99.382	102.964	107.419	115.770	120.120	126.082
125	97.431	100.021	104.028	107.713	112.300	120.916	125.412	131.583
130	101.911	104.569	108.684	112.470	117.189	126.066	130.708	137.087
135	106.402	109.128	113.349	117.236	122.084	131.221	136.006	142.592
140	110.904	113.697	118.023	122.009	126.984	136.379	141.308	148.100
145	115.417	118.276	122.706	126.789	131.891	141.541	146.613	153.610
150	119.940	122.864	127.396	131.575	136.803	146.706	151.920	159.122
155	124.473	127.461	132.093	136.368	141.720	151.874	157.230	164.636
160	129.014	132.065	136.797	141.167	146.641	157.046	162.542	170.151
165	133.564	136.678	141.508	145.972	151.567	162.220	167.856	175.668
170	138.123	141.298	146.225	150.781	156.498	167.397	173.172	181.187
175	142.688	145.924	150.948	155.595	161.432	172.576	178.490	186.706
180	147.262	150.558	155.677	160.415	166.370	177.758	183.810	192.227
185	151.842	155.198	160.411	165.239	171.312	182.942	189.133	197.750
190	156.429	159.844	165.151	170.068	176.257	188.129	194.456	203.273
195	161.023	164.496	169.895	174.901	181.206	193.318	199.781	208.797
200	165.623	169.154	174.644	179.738	186.161	198.508	205.108	214.323

Table 5 *Offered Traffic Load in Erlangs in an Erlang-C System (Number of Channels from 2 to 36)*

Channel Number	Probability of Non-Zero Delay								
	0.01	**0.02**	**0.05**	**0.07**	**0.10**	**0.20**	**0.50**	**0.70**	**0.100**
2	0.147	0.210	0.342	0.411	0.500	0.740	1.281	1.584	2.000
3	0.429	0.554	0.787	0.900	1.040	1.393	2.116	2.496	3.000
4	0.810	0.994	1.319	1.469	1.653	2.102	2.977	3.422	4.000
5	1.259	1.497	1.905	2.090	2.313	2.847	3.856	4.357	5.000
6	1.758	2.047	2.532	2.748	3.007	3.617	4.747	5.299	6.000
7	2.296	2.633	3.188	3.434	3.725	4.406	5.646	6.245	7.000
8	2.866	3.246	3.869	4.141	4.463	5.210	6.553	7.195	8.000
9	3.460	3.883	4.569	4.867	5.218	6.027	7.466	8.149	9.000
10	4.077	4.540	5.285	5.607	5.986	6.853	8.383	9.104	10.000
11	4.712	5.213	6.015	6.361	6.765	7.688	9.304	10.062	11.000
12	5.362	5.901	6.758	7.125	7.554	8.530	10.229	11.022	12.000
13	6.027	6.601	7.511	7.899	8.352	9.379	11.157	11.983	13.000
14	6.705	7.313	8.273	8.682	9.158	10.233	12.088	12.946	14.000
15	7.394	8.035	9.044	9.473	9.970	11.093	13.021	13.911	15.000
16	8.093	8.766	9.822	10.270	10.789	11.958	13.957	14.876	16.000
17	8.801	9.505	10.607	11.074	11.613	12.826	14.894	15.842	17.000
18	9.517	10.252	11.399	11.883	12.443	13.699	15.834	16.810	18.000
19	10.242	11.006	12.196	12.698	13.277	14.575	16.774	17.778	19.000
20	10.973	11.766	12.998	13.517	14.116	15.454	17.717	18.747	20.000
21	11.711	12.532	13.806	14.341	14.958	16.336	18.661	19.718	21.000
22	12.455	13.304	14.618	15.169	15.805	17.221	19.606	20.688	22.000
23	13.205	14.081	15.434	16.001	16.654	18.109	20.553	21.659	23.000
24	13.960	14.862	16.254	16.837	17.508	18.999	21.500	22.631	24.000
25	14.721	15.648	17.078	17.676	18.364	19.892	22.449	23.604	25.000
26	15.486	16.438	17.905	18.518	19.223	20.786	23.399	24.577	26.000
27	16.255	17.233	18.736	19.364	20.084	21.683	24.350	25.551	27.000
28	17.029	18.031	19.569	20.212	20.948	22.581	25.301	26.525	28.000
29	17.806	18.832	20.406	21.062	21.815	23.481	26.254	27.499	29.000
30	18.588	19.637	21.246	21.916	22.684	24.383	27.207	28.474	30.000
31	19.373	20.446	22.088	22.772	23.555	25.287	28.161	29.450	31.000
32	20.162	21.257	22.932	23.630	24.428	26.192	29.116	30.425	32.000
33	20.953	22.071	23.780	24.490	25.303	27.099	30.071	31.401	33.000
34	21.748	22.888	24.629	25.353	26.180	28.007	31.028	32.378	34.000
35	22.546	23.708	25.481	26.217	27.059	28.916	31.984	33.355	35.000
36	23.347	24.530	26.335	27.084	27.940	29.827	32.942	34.332	36.000

Table 6 *Offered Traffic Load in Erlangs in an Erlang-C System (Number of Channels from 37 to 71)*

Channel Number	Probability of Non-Zero Delay								
	0.01	0.02	0.05	0.07	0.10	0.20	0.50	0.70	0.100
37	24.151	25.355	27.190	27.952	28.822	30.739	33.900	35.309	37.000
38	24.957	26.182	28.048	28.822	29.706	31.652	34.859	36.287	38.000
39	25.765	27.011	28.908	29.694	30.591	32.566	35.818	37.265	39.000
40	26.577	27.843	29.769	30.567	31.478	33.482	36.777	38.244	40.000
41	27.390	28.676	30.632	31.442	32.366	34.398	37.738	39.222	41.000
42	28.206	29.512	31.497	32.318	33.256	35.316	38.698	40.201	42.000
43	29.024	30.350	32.363	33.196	34.147	36.234	39.659	41.181	43.000
44	29.844	31.189	33.231	34.076	35.039	37.153	40.621	42.160	44.000
45	30.666	32.030	34.101	34.956	35.932	38.074	41.583	43.140	45.000
46	31.490	32.873	34.972	35.838	36.827	38.995	42.545	44.119	46.000
47	32.316	33.718	35.844	36.722	37.722	39.917	43.508	45.099	47.000
48	33.143	34.564	36.717	37.606	38.619	40.840	44.471	46.080	48.000
49	33.973	35.412	37.592	38.492	39.517	41.763	45.435	47.060	49.000
50	34.804	36.262	38.469	39.379	40.416	42.688	46.399	48.041	50.000
51	35.637	37.113	39.346	40.267	41.316	43.613	47.363	49.022	51.000
52	36.471	37.965	40.225	41.156	42.217	44.539	48.328	50.003	52.000
53	37.308	38.819	41.104	42.046	43.119	45.466	49.293	50.984	53.000
54	38.145	39.674	41.985	42.938	44.021	46.393	50.258	51.965	54.000
55	38.985	40.531	42.867	43.830	44.925	47.321	51.224	52.947	55.000
56	39.825	41.389	43.750	44.723	45.830	48.250	52.190	53.929	56.000
57	40.667	42.248	44.635	45.617	46.735	49.179	53.156	54.911	57.000
58	41.511	43.108	45.520	46.512	47.641	50.109	54.123	55.893	58.000
59	42.355	43.970	46.406	47.408	48.548	51.039	55.090	56.875	59.000
60	43.202	44.833	47.293	48.305	49.456	51.970	56.057	57.858	60.000
61	44.049	45.696	48.181	49.203	50.365	52.902	57.024	58.840	61.000
62	44.897	46.561	49.070	50.102	51.274	53.834	57.992	59.823	62.000
63	45.747	47.427	49.960	51.001	52.184	54.767	58.960	60.805	63.000
64	46.598	48.295	50.851	51.901	53.095	55.700	59.928	61.788	64.000
65	47.450	49.163	51.742	52.802	54.006	56.634	60.897	62.772	65.000
66	48.304	50.032	52.635	53.704	54.918	57.568	61.865	63.755	66.000
67	49.158	50.902	53.528	54.606	55.831	58.503	62.834	64.738	67.000
68	50.014	51.773	54.422	55.510	56.745	59.439	63.804	65.722	68.000
69	50.870	52.645	55.317	56.413	57.659	60.374	64.773	66.705	69.000
70	51.728	53.518	56.212	57.318	58.574	61.311	65.743	67.689	70.000
71	52.586	54.392	57.108	58.223	59.489	62.247	66.712	68.673	71.000

Table 7 *Offered Traffic Load in Erlangs in an Erlang-C System (Number of Channels from 72 to 130)*

Channel Number	Probability of Non-Zero Delay								
	0.01	0.02	0.05	0.07	0.10	0.20	0.50	0.70	0.100
72	53.446	55.267	58.006	59.129	60.405	63.184	67.683	69.657	72.000
73	54.306	56.143	58.903	60.036	61.321	64.122	68.653	70.641	73.000
74	55.168	57.019	59.802	60.943	62.238	65.060	69.623	71.625	74.000
75	56.030	57.897	60.701	61.851	63.156	65.998	70.594	72.609	75.000
76	56.894	58.775	61.601	62.759	64.074	66.937	71.565	73.594	76.000
77	57.758	59.654	62.501	63.668	64.993	67.876	72.536	74.578	77.000
78	58.623	60.533	63.402	64.578	65.912	68.816	73.507	75.562	78.000
79	59.489	61.414	64.304	65.488	66.832	69.756	74.478	76.547	79.000
80	60.356	62.295	65.206	66.399	67.752	70.697	75.450	77.532	80.000
81	61.224	63.177	66.109	67.311	68.673	71.637	76.422	78.517	81.000
82	62.092	64.060	67.013	68.223	69.594	72.579	77.394	79.502	82.000
83	62.961	64.943	67.917	69.135	70.516	73.520	78.366	80.487	83.000
84	63.831	65.827	68.822	70.048	71.438	74.462	79.338	81.472	84.000
85	64.702	66.712	69.727	70.961	72.361	75.404	80.311	82.457	85.000
86	65.574	67.598	70.633	71.876	73.284	76.347	81.283	83.443	86.000
87	66.446	68.484	71.539	72.790	74.208	77.290	82.256	84.428	87.000
88	67.319	69.371	72.446	73.705	75.132	78.233	83.229	85.413	88.000
89	68.193	70.258	73.354	74.621	76.056	79.176	84.202	86.399	89.000
90	69.067	71.146	74.262	75.536	76.981	80.120	85.175	87.385	90.000
91	69.943	72.035	75.170	76.453	77.906	81.064	86.149	88.370	91.000
92	70.818	72.924	76.079	77.370	78.832	82.009	87.122	89.356	92.000
93	71.695	73.814	76.989	78.287	79.758	82.954	88.096	90.342	93.000
94	72.572	74.705	77.899	79.205	80.685	83.898	89.070	91.328	94.000
95	73.450	75.596	78.809	80.123	81.612	84.844	90.044	92.314	95.000
96	74.329	76.488	79.720	81.042	82.539	85.790	91.018	93.300	96.000
97	75.208	77.380	80.631	81.961	83.466	86.735	91.992	94.286	97.000
98	76.087	78.273	81.543	82.880	84.394	87.682	92.966	95.273	98.000
99	76.968	79.166	82.455	83.800	85.323	88.628	93.941	96.259	99.000
100	77.849	80.060	83.368	84.720	86.252	89.575	94.915	97.245	100.000
105	82.262	84.537	87.938	89.328	90.901	94.313	99.790	102.178	105.000
110	86.690	89.026	92.519	93.944	95.558	99.056	104.668	107.113	110.000
115	91.130	93.527	97.108	98.570	100.223	103.806	109.549	112.049	115.000
120	95.582	98.039	101.707	103.203	104.895	108.561	114.432	116.986	120.000
125	100.046	102.561	106.314	107.844	109.574	113.321	119.318	121.925	125.000
130	104.520	107.092	110.928	112.492	114.259	118.086	124.206	126.865	130.000

Table 8 *Offered Traffic Load in Erlangs in an Erlang-C System (Number of Channels from 135 to 200)*

Channel Number	Probability of Non-Zero Delay								
	0.01	0.02	0.05	0.07	0.10	0.20	0.50	0.70	0.100
135	109.004	111.632	115.550	117.146	118.951	122.855	129.096	131.805	135.000
140	113.498	116.181	120.179	121.808	123.648	127.629	133.988	136.748	140.000
145	118.007	120.751	124.848	126.520	128.413	132.516	139.047	141.827	145.000
150	122.518	125.315	129.490	131.193	133.120	137.297	143.943	146.771	150.000
155	127.038	129.887	134.137	135.871	137.833	142.082	148.840	151.716	155.000
160	131.565	134.465	138.791	140.555	142.550	146.871	153.740	156.662	160.000
165	136.099	139.050	143.449	145.243	147.271	151.663	158.640	161.608	165.000
170	140.640	143.641	148.113	149.936	151.997	156.458	163.543	166.556	170.000
175	145.180	148.223	152.748	154.587	156.663	161.145	168.280	171.368	175.000
180	149.735	152.826	157.421	159.289	161.396	165.945	173.185	176.317	180.000
185	154.295	157.434	162.099	163.994	166.133	170.749	178.091	181.267	185.000
190	158.862	162.048	166.781	168.704	170.874	175.555	182.999	186.217	190.000
195	163.434	166.666	171.468	173.418	175.618	180.364	187.908	191.168	195.000
200	168.011	171.290	176.158	178.135	180.365	185.175	192.818	196.120	200.000

Tables

Appendix-D

Table 1 *Tabulation of the Q-function*

z	Q(z)	z	Q(z)
0.0	0.50000	2.0	0.02275
0.1	0.46017	2.1	0.01786
0.2	0.42074	2.2	0.01390
0.3	0.38209	2.3	0.01072
0.4	0.34458	2.4	0.00820
0.5	0.30854	2.5	0.00621
0.6	0.27425	2.6	0.00466
0.7	0.24196	2.7	0.00347
0.8	0.21186	2.8	0.00256
0.9	0.18406	2.9	0.00187
1.0	0.15866	3.0	0.00135
1.1	0.13567	3.1	0.00097
1.2	0.11507	3.2	0.00069
1.3	0.09680	3.3	0.00048
1.4	0.08076	3.4	0.00034
1.5	0.06681	3.5	0.00023
1.6	0.05480	3.6	0.00016
1.7	0.04457	3.7	0.00011
1.8	0.03593	3.8	0.00007
1.9	0.02872	3.9	0.00005

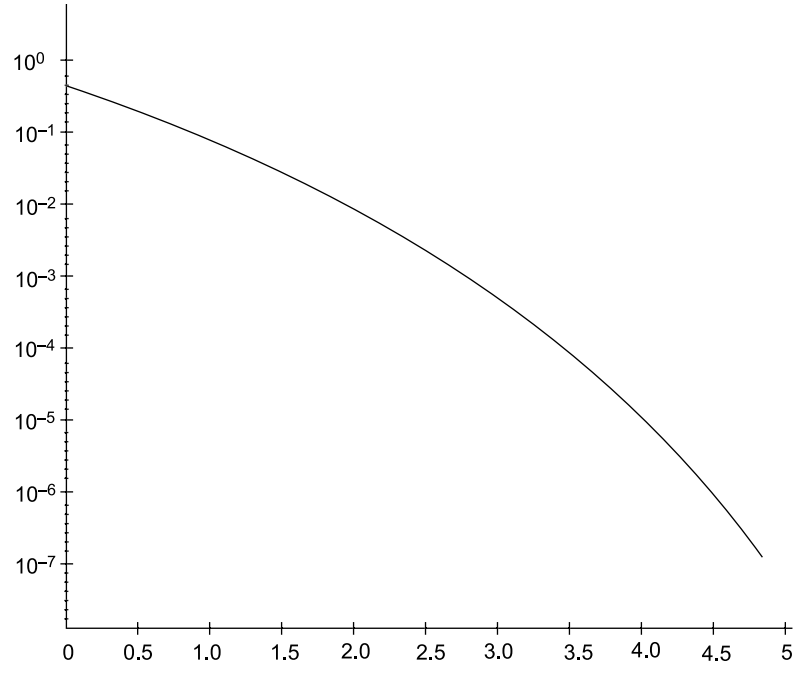


Fig. C.1 *Plot of the Q-function*

erf and erfc Functions

The error function (*erf*) is defined as

$$erf(z) = \frac{2}{\sqrt{\pi_0}} \int_0^z e^{-x^2} dx$$

and the complementary error function (*erfc*) is defined as

$$erfc(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-x^2} dx$$

The *erfc* function is related to the *erf* function by

$$erfc(z) = 1 - erf(z)$$

The *Q*-function is related to the *erf* and *erfc* functions by

$$Q(z) = \frac{1}{2}\left[1 - erf\left(\frac{z}{\sqrt{2}}\right)\right] = \frac{1}{2} erfc\left(\frac{z}{\sqrt{2}}\right)$$

$$erfc(z) = 2Q(\sqrt{2z})$$

$$erf(z) = 1 - 2Q(\sqrt{2z})$$

Table 2 *Tabulation of the Error Function erf(z)*

z	erf(z)	z	erf(z)
0.1	0.11246	1.6	0.97635
0.2	0.22270	1.7	0.98379
0.3	0.32863	1.8	0.98909
0.4	0.42839	1.9	0.99279
0.5	0.52049	2.0	0.99532
0.6	0.60385	2.1	0.99702
0.7	0.67780	2.2	0.99814
0.8	0.74210	2.3	0.99885
0.9	0.79691	2.4	0.99931
1.0	0.84270	2.5	0.99959
1.1	0.88021	2.6	0.99976
1.2	0.91031	2.7	0.99987
1.3	0.93401	2.8	0.99993
1.4	0.95228	2.9	0.99996
1.5	0.96611	3.0	0.99998

Question Papers

Appendix E

SAMPLE PAPER 1: MOBILE AND WIRELESS COMMUNICATIONS

Each question carries 20 marks

Q.1 (i) Discuss the challenges that a wireless channel faces.
(ii) How do time dispersion and fading occur in wireless environments?
(iii) Show the amplitude variation of a signal in a wireless media with two propagation paths.
(iv) How does frequency dispersion occur due to Doppler effect in a wireless channel?

Q.2 (i) How is the number of a cell related with cluster size? Derive the expression.
(ii) Consider a cellular system with *S/I* ratio of 19 dB. The frequency reuse factor is $N = 7$, calculate the worst case for signal-to-co-channel interference ratio. Is the frequency reuse factor 7 still being acceptable? If not, what is it?
(iii) How is cell splitting useful to increase the capacity of cellular systems? Explain with illustration.

Q.3 (i) Highlight the significant advancement of mobile services offered by wireless networks from 1G to 3G-evolution path.
(ii) What does GSM stand for? Write the operating frequency of GSM for uplink and downlink channels?
(iii) Tabulate the different standards for 2G.
(iv) Define the following terms with their functionalities:
IMSI, IMEI, T-IMSI
(v) Describe the GSM call set-up procedure.

Q.4 (i) Layout the GPRS interfaces. Explain the function of *Gn*, *Gp* and *Gi* interfaces.
(ii) Give the GPRS protocol stacks for user plane and control plane.
(iii) Highlight the functionalities for SNDCP and BSSGP protocols.

Q.5 (i) What do you mean by PDP context for GPRS network?
(ii) How does data transfer occur through a GPRS network?
(iii) What is the role of GRX (GPRS roaming exchange)?

Q.6 (i) How does path loss change in case of two-ray model with respect to free space model?
(ii) Describe Okumara–Hata model.
(iii) Obtain the expression for received signal for the transmitting signal $x(t)$ considering Doppler effect.
(iv) Differentiate frequency selective fading with time dispersion and frequency dispersion with time selective fading.

Q.7 (i) Explain the principle of operation of PHY layer for IEEE 802.11 standard.
(ii) What are the families of IEEE 802.11? Mention their main features.
(iii) Define the terms BSS and ESS in respect of IEEE 802.11.

Q.8 (i) What are the two main integration models for WLAN with cellular networks?
(ii) Describe the main benefits of the integration architecture. What are the main constraints arising due to integration?

SAMPLE PAPER 2: MOBILE AND WIRELESS COMMUNICATIONS

Each question carries 20 marks

Q.1 (i) Define the terms 'cell' and "cluster" for cellular networks. What is the importance of cell clustering for cellular communications?

(ii) How is frequency reuse helpful in capacity expansion? Explain with an example.

(iii) How are locations of co-channel cells determined in a cellular system? Explain with pictorial representation.

Q.2 (i) Establish the relationship between the co-channel cell distance (D) and the cell radius R for a hexagonal cellular structure.

(ii) What is co-channel interference? How is signal to interference ratio (S/I) related with frequency reuse ratio q?

(iii) Discuss the concept of macro, micro and pico cells. When do they become useful?

(iv) If a mobile is located at the cell boundary and experiences worst-case co-channel interference on the forward channel, how does S/I ratio change? Explain this situation with respect to $N = 7$ (frequency reuse factor).

Q.3 (i) Discuss the various method of increasing capacity in a cellular wireless system? Highlight the pros and cons of those methods.

(ii) When a cell with radius R is split to a new cell of radius $R/2$, what will be the base station transmit power for the two cases. Find the transmit power ratio in dB.

(iii) Consider a cellular system with hexagonal cells of radius R and cell cluster size 7, the base station transmit power cannot satisfy the 18 dB S/I ratio requirements. Determine whether use of 120^o sectoring can satisfy 18 dB requirements with same 7 cell frequency reuse. What will happen for 60^o sectoring?

Q.4 (i) Consider a cellular system in which there are a total of 1000 radio channels available for traffic handling. The area of the cell is 5 km^2 and entire area is 2000 km^2. Calculate the system capacity with $N = 7$ and 4. Find if decreasing of cluster size increases system capacity and with what capacity for this particular problem?

(ii) What is the role of S/I ratio in the design of cell cluster?

Q.5 (i) What are the main characteristics of AMPS networks?

(ii) List different 1G systems available with their main features.

(iii) What are the limitations of 1G cellular networks?

(iv) Digital technology is the breakthrough for 2G cellular systems—explain.

(v) What are the main differences in 2G and 3G cellular systems?

Q.6 (i) How is GPRS network evolved from GSM network? What is the other 2.5G variant of GSM?

(ii) Discuss the main objectives of IMT-2000. What are the two main project branches under IMT2000?

(iii) Discuss the role of two new nodes in GPRS for packet data service.

Q.7 (i) Explain the activation of PDP context in GPRS network.

(ii) Discuss the transparent and non-transparent methods of data transport in GPRS networks.

Q.8 (i) Explain the operation of DCF and PCF in IEEE 802.11 standard.

(ii) What are the necessity of RTS and CTS commands?

(iii) What is the role of NAV in WLAN system?

(iv) How is CSMA/CA protocol implemented in WLAN with back window? Explain with a circuit diagram.

SAMPLE PAPER 3: MOBILE AND WIRELESS COMMUNICATIONS

Each questions carries 20 marks

Q.1 (i) What are the two main functions of mobility management? How is location tracked within a cellular network (GSM) while a user roaming?
(ii) Describe the procedures for call set-up in GSM networks.
(iii) How is data tunneled to and from GGSN using GTP?

Q.2 (i) What are the new interfaces defined in UMTS network? Will a mobile in GPRS network operate in UMTS? Give reasons for your answer.
(ii) What major changes need to occur in UMTS release 99 from GPRS?
(iii) What is the access technology for 3G UMTS network?

Q.3 (i) Describe the IEEE 802.11 architecture.
(ii) Explain DSS and FSS techniques.
(iii) Why cannot WLAN implement CSMA/CD?

Q.4 (i) WiMAX is the revolution in the next generation wireless broadband communication—comment on this issue.
(ii) Explain the operation of adaptive modulation and coding in WiMAX PHY.
(iii) Throughput capacities of WiMAX is dependent on channel bandwidth. How?

Q.5 (i) Discuss the major roles of the following entities with respect to GSM network architecture.
- Mobile switching center (MSC)
- Home Location Register (HLR)
- Authentication center (AC)

(ii) Explain the operation of multiple access and duplexing technique for GSM.
(iii) Discuss about the logical channel structure in GSM.

Q.6 (i) A cellular network consists of radio cells. What does a radio cell mean? On what factors does the coverage of a cell depend?
(ii) Associated with each capacity enhancement is a price tag. What are the prices to be paid for the different capacity enhancement in cellular communication?
(iii) What is the significance of the numbers IMSI and EIR with respect to cellular mobile communication?

Q.7 (i) Discuss the problem created due to user roaming, and the mechanisms needed to ensure seamless information delivery to the mobile user at its current location.
(ii) Explain the triangular routing in mobile IP operation. How can this triangular routing be eliminated?
(iii) What are the limitations of Mobile IP? How are these overcome?

Q.8 (i) Derive the expression for (S/I) ratio in a worst-case scenario with 60^{o} sectorization.
(ii) Prove that $(S/I)_{120}/ (S/I)_{omni} = 3$
(iii) Determine the cell cluster size N, with respect to two integers i and j association in determining co-channel cells.

Q.9 (i) Integration of cellular-WLAN architectures is the prospective opportunities to both the users and operators. Discuss this issue.
(ii) There are many technological challenges to be solved to integrate the heterogeneous architecture. Point out some of the problems that will arise due to integration.

SAMPLE PAPER 4: MOBILE AND WIRELESS COMMUNICATIONS

Each question carries 20 marks

Q.1 (a) Explain the operational techniques of TDMA. Give the TDMA frame structure.

(b) Discuss the pros and cons of TDMA techniques over FDMA.

(c) With the pictorial representation explain the use of TDMA/FDD in GSM systems.

(d) Total bandwidth in AMPS cellular system is allocated 12.5 MHz. Using FDMA, 416 numbers of available channels with spacing 30 KHz is allocated to the users.

(i) What is the guard bandwidth used in the system?

(ii) What is the spectral efficiency for this system if there are 21 channels used for control signaling?

(iii) If the cell area is 8 km^2 and frequency reuse factor is 4, find the overall spectral efficiency of the system.

(iv) If the trunk efficiency of the system is 0.9, what is the spectral efficiency in Earlangs/MHz/km^2

Q.2 (a) What are the different mechanisms related to multipath propagation.

(b) Let Ws be the signal bandwidth and T_m be the multipath delay spread of the channel. Now differentiate the four cases of multipath effects in mobile wireless communication with respect to the conditions related to W_s and T_m.
Frequency selective and time selective fading and
Frequency dispersion and time dispersion fading

(c) A WSSUS channel has a multipath delay spread T_m =1 s and A Doppler spread of f_d = 0.02 Hz. The total channel bandwidth at band pass available for signal transmission is W_s = 5 Hz. To reduce the effects of ISI, the signal designer selects a symbol duration T_s =10s

(i) Determine the coherence bandwidth and the coherence time

(ii) Does the channel exhibit frequency selective fading? Explain.

(iii) Does this channel exhibit slow and fast fading? Explain.

(iv) Determine the transmission data rate of the system.

(d) Establish the relation between time-variant transfer function H (f, t) and the Doppler spread function H(f,γ)?

Q.3 (a) In the radio cell layout, in addition to the hexagonal topology, a square or an equilateral triangle topology can also be used.
Given the same distance between the cell centre and its furthest perimeter points as R, compare the cell coverage areas among the three regular polygons (hexagon, square and triangle). Discuss the advantages of using the hexagonal cell shape over the square and triangle cell shapes

(b) A cellular service provider decides to use a digital TDMA scheme which can tolerate a signal-to-interference ratio of 15 dB in the worst case. Find the optimal value of N for (a) omni-directional antennas, (b) 120° sectoring, and (c) 60° sectoring. Should sectoring be used? If so, which case (60° or 120°) should be used? (Assume a path loss exponent of k = 4 and consider the effect of trunking efficiency).

(c) Compute the signal to interference (SIR) ration at the mobile when only, the first tier are taken into consideration for the interfering cells. Assume power received at a distance d from the centrally placed base station, with omnidirectional antenna is $P_r = P_t\left(\frac{1}{d}\right)^k$.
Give a comparative results for cluster size N=1, 3, 4 and 7

Q.4 (a) (i) At what frequency GSM base station start transmitting?
(ii) Why does mobile station transmit at lower frequency?
(iii) One time slot of TDMA frame is
Logical Channel / Physical channel
(iv) Personal mobility can be supported because of SIM/HLR/VLR
(v) The main difference between the GSM SIM and USIM is that
A. USIM is downloadable and can be accessed via air interface and modified by the network
B. It can store JAVA application
C. Higher capacity than GSM SIM
D. All of the above

(b) How is data transferred to external PDN through GPRS network?
(c) What are the main functions of SNDCP protocol?
(d) Describe the GSM frame structures.

Q.5 (a) Define hand-off process and it's type. When is priority of handoff call used?
(b) Distinguish log-normal shadowing and log-distance path loss model.
(c) The discrete power profile delay for multipath transmission is shown in figure below, show the multipath power gain, mean delay and rms delay spread.

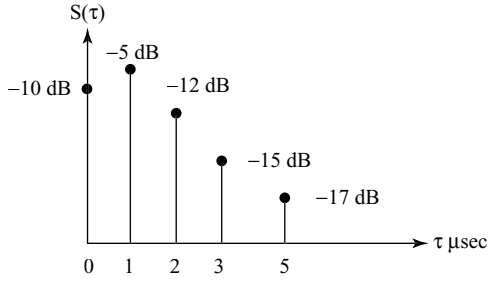


Fig. S4

(d) What are the different types of channel assignment schemes used for cellular communication?

Q.6 Write short notes on any four of the following:
(a) Adaptive Equalization process
(b) OFDM access
(c) Spectral efficiency for CDMA system
(d) Location update of GSM network
(e) PDP context activation of GPRS network

Q.7 (a) What are the architectural changes required to support packet data over GSM networks?
(b) Discuss the process of PDP context activation in GPRS network. What is the relation between PDP context and session management?
(c) Explain how is mobility of a user tracked in GPRS system?

Q.8 (a) Show the evolution processes from 1G to 3G networks with main technological focuses in terms of data rate and application.
(b) Why is CDMA based cellular system broadband?
(c) What was the main objective of 3GPP and 3GPP2 projects?
(d) Discuss the evolutionary changes occur in the 3G UMTS networks from release 99 to release 5.

SAMPLE PAPER 5: MOBILE AND WIRELESS COMMUNICATIONS

Each question carries 20 marks

Q.1 (a) Describe the possible cell structures in a cellular system.
(b) Why do we require cell cluster?
(c) Determine the cell cluster size N in a hexagonal cellular array.
(d) What is called frequency reuse ratio "q"?
(e) Determine the S/I ratio in terms of "q"?
A cellular telephone system with 110 channels uses a modulation scheme requiring a minimum S/I ratio of 19 dB for acceptable link performance. Assume that the propagation loss is only distance dependent and increases with the fourth power of the distance. Determine channels/cell that can be offered by the system. Assume the hexagonal cell structure with base station at the center and transmitting same power.

Q.2 (a) Why does propagation path loss model require in wireless communications?
(b) With respect to the pictorial representation, mention different path loss models dependant on receiver antenna location.
(c) Derive the expression for EIRP (Effective Isotropic Radiated Power). How is it related with ERP?
(d) Differentiate the operations of time, frequency dispersion and time, frequency selective fading.
(e) If the received power at a reference distance d_0= 1 km is equal to 1 mW, find the received power at distances of 2 km, 4 km and 8 km from the same transmitter using free space path loss model and with k = 4. Comment on the results.

Q.3 (a) If a cell is sectored into 120°, what will be the improvement of S/I ratio over omni-directional cell?
(b) What are the three mutually independent multiplicative propagation phenomena?
(c) The large-scale effects determine a power level averaged over _____where as small-scale effects vary a received power over _____.
(d) In the context of cell splitting, the _____ is known as pico-cell, the _____ is the micro-cell and the _____ without splitting is called the macro-cell.
(e) The _____ is the mostly used macro-cell coverage planning. The coverage area ranges over _____.
(f) For _____ model the average received signal power decreases logarithmically with _____ whether in indoor or an outdoor environment.
(g) When the signal with _____ delays recombine, they produce distortion known as _____ fading. But when recombination of _____ delays causes _____ fading.
(h) The _____ distribution is a well-known statistic for the amplitude modeling of radio signals in a _____ environments.
(i) The motion of the mobile receiver causes _____ in _____ domain. It causes the spectral _____.
(j) _____ antennas are generally used in base station site whereas for mobile handset _____ antenna is very common.
(k) The billing in GPRS system id done by the amount of _____, rather than _____ time.
(l) The uplink and downlink frequency range for AMPS system are _____ and _____ respectively.

(m) WCDMA uses _____ and cdma2000 uses _____ techniques.
(n) _____ is the main modulation scheme used in 4G systems.
(o) The word MAGIC stands for _____, _____, _____, _____, _____ is just appropriate to defines 4G.

Q.4 (a) Explain the operational techniques of TDMA. Give the TDMA frame structure.
(b) Discuss the pros and cons of TDMA techniques over FDMA.
(c) With the pictorial representation explain the use of TDMA/FDD in GSM systems.
(d) Total bandwidth in AMPS cellular system is allocated 12.5 MHz. Using FDMA, 416 numbers of available channels with spacing 30 KHz is allocated to the users.
(e) What is the guard bandwidth used in the system?
(f) What is the spectral efficiency for this system if there are 21 channels used for control signaling?
(g) If the cell area is 8 km^2 and frequency reuse factor is 4, find the overall spectral efficiency of the system.
(h) If the trunk efficiency of the system is 0.9, what is the spectral efficiency in Earlangs/MHz/km^2.

Q.5 (a) CDMA system is different from FDMA/TDMA systems. For getting spectral efficiency, the term E_b/N_0 (bit energy per total noise power spectral density) plays a great role for DS-CDMA. By defining suitable terms, calculate the number of users that can be supported in a DS-CDMA system to define its spectral efficiency.
(b) A CDMA system is defined with the following parameters:
Frequency reuse efficiency $\eta_f = 0.65$, $E_b/ N_0 = 12$ dB, the information bit transmission rate is 19.2 kbps, system bandwidth is W= 12.5 MHz. Neglecting all other sources of interference determine the system capacity and spectral efficiency of the CDMA system.
(c) Authentication is the process to prove the identity of valid users claiming services of the network – how is authentication checked in GSM cellular systems?

Q.6 (a) Discuss the functionalities of the following entities in 2G GSM networks: BSC, MSC, HLR and VLR.
(b) What are architectural changes required to support packet data over GSM networks?
(c) Discuss the process of PDP context activation in GPRS network. What is the relation between PDP context and session management?

Q.7 (a) Discuss the forward and reverse link characteristics in IS-95 CDMA systems?
(b) Why is power control important in CDMA networks? How many types of power control mechanisms are there?
(c) How is base station identified by the Mobile handset in CDMA system?

Q.8 (a) What are the main operational steps in Mobile IPv4? Discuss the limitations of MIPv4.
(b) Discuss the transparent and non-transparent access of packet data transfer over GPRS network?

Glossary of Terms and Abbreviations

1G	First Generation
2G	Second Generation
3G	Third Generation
3GPP	Third Generation Partnership Project
3GPP2	Third Generation Partnership Project 2
AAA	Authorization, Authentication and Accounting
ACI	Adjacent Channel Interference
AGCH	Access Grant Channel
AGUA	Aggregateable Global Unicast Address
ALCAP	Access Link Control Application Protocol
ALL2	ATM Adaptation Layer 2
ALL5	ATM Adaptation Layer 5
AKA	Authentication and Key Agreement
AMPS	Advanced Mobile Phone Systems
APN	Access Point Name
ARP	Address Resolution Protocol
ARQ	Automatic Repeat Request
AS	Autonomous Systems
ASK	Amplitude Shift Keying
ASN	Access Service Network
ATM	Asynchronous Transfer Mode
ATDMA	Asynchronous Time Division Multiple Access
AuC	Authentication Centre
AUTN	Authentication Token
AWGN	Additive White Gaussian Noise
BCCH	Broadcast Control Channel
BER	Bit Error Rate
BICC	Bearer Independent Call Control
BG	Boarder Gateway
BGP	Boarder Gateway Protocol
BMC	Broadcast/Multicast Control
BPSK	Binary Phase Shift Keying
BS	Base Station
BSSAP	Base System Application Part
BSSGP	Base Station Subsystem GPRS Protocol
BVCI	BSSGP Virtual Connection (BVC) Identifier
BSC	Base Station Controller
BSIC	Base Station Identity Code
BSS	Basic Service Set
BSS-ID	Basic Service Set Identifier
BSS	Base Station Subscriber
BTS	Base Transceiver Station

BTS-TRX	Base Station Transceiver
BU	Binding Update
BWA	Broadband Wireless Access
CBS	Customer Care and Billing System
CCA	Clear Channel Assessment
CCCH	Common Control Channel
CCITT	Comtte Consultatif International Telegraphique et Telephonique (The International Telegraph and Telephone Consultative Committee)
CCK	Complementary Code Keying
CCS	Common Channel Signaling
CDMA	Code Division Multiple Access
CDMA2000	Code Division Multiple Access 2000
CDG	CDMA Development Group
CGW/CCF	Charging Gateway/Charging Collection Function
C/I	Carrier-to-Interference
Ch	Channel
CH	Corresponding Host
CER	Customer Edge Router
CFP	Contention Free Period
CG	Charging Gateway
CGI	Cell Global Identity
CLNS	Connectionless Network Service
CLPC	Close Loop Power Control
CLL-TE	Common LLC Layer Terminal Equipment
CIP-TE	Common IP Layer Terminal Equipment
CM	Connection Management
CMM	Common Mobility Management
CN	Core network
CONS	Connection Oriented Network Service
CoA	Care-of Address
CP	Contention Period
CPCH	Common Packet Channel
CRC	Cyclic Redundancy Check
CRNC	Controlling RNC
CS	Circuit switching
CSCF	Call State Control Function
CSMA	Carrier Sense Multiple Access
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CSMA/CD	Carrier Sense Multiple Access with Collision Detection
CSN	Connectivity Service Network
CTIA	Cellular Telecommunications and Internet Association
CTS	Clear To Send
CT2	Cordless Telephone
CW	Contention Window
DARS	Digital Audio-Radio Services
DCA	Dynamic Channel Assignment
DCD	Downlink Channel Descriptor
DCF	Distributed Coordination Function

DCS	Digital Cellular System
DCCH	Dedicated Control Channel
DCCP	Datagram Congestion Control Protocol
D-DP	Delay Doppler Spread
DDNS	Dynamic Domain Name System
DECT	Digital European Cordless Telecommunications
DFE	Decision Feedback Equalizer
DFS	Dynamic Frequency Selection
DHAAD	Dynamic Home Agent Address Discovery
DHCP	Dynamic Host Configuration Protocol
DIFS	Distributed Coordination Function (DCF) Inter frame Space
DISN	Destination Initial Sequential Number
DIUC	Downlink Interval Usage Code
DI-MAP	Downlink MAP
DLC	Data Link Control
DLCI	Data Link Connection Identifier
DLL	Data Link Layer
DNS	Domain Name System
DPDCH	Dedicated Physical Channel
DPCCH	Dedicated Physical Control Channel
DQPSK	Differential Quadrature Phase Shift Keying
DRNC	Drift RNC
DS	Distributed System
DS-CDMA	Direct Sequence Code Division Multiple Access
DSF	Doppler Spread Function
DSL	Digital Subscriber Line
DSS	Distribution System Service
DSSS	Direct Sequence Spread Spectrum
DTAP	Direct Transfer Application Part
DTCH	Dedicated Traffic Channel
D/U	Desired to Undesired Signal
DV	Data and Voice
EAP	Extensible Authentication Protocol
EAPOL	Extensible Authentication Protocol (EAP) over LAN
EDCF	Enhanced Distributed Coordination Function
EDGE	Enhanced Data Rates for Global GSM Evolution
EGPRS	Enhanced GPRS System
EIR	Equipment Identity Register
EIRP	Effective Isotropic Radiated Power
EKS	Encryption Key Sequence
ERP	Effective Radiated Power
ESS	Extended Service Set
ESS-ID	ESS Identifier
ETSI	European Telecommunications Standard Institute
FA	Foreign Agent
FACoA	Foreign Agent Care-of- Address
FACCH	Fast Associated Control Channel

FBCCh	Forward Broadcast Control Channels
FBI	Feedback Information
FC	Frame Control
FCA	Fixed Channel Assignment
FCC	Federal Communications Commission
FCCH	Frequency Correction Channel
FCCh	Forward Common Control Channels
FCH	Frequency Channel
FCS	Forward Check Sequence
FEC	Forward Error Checking
FPiCh	Forward Pilot Channel
FPCh	Forward Paging Channel
FDD	Frequency Division Duplex
FDM	Frequency Division Multiplexing
FDMA	Frequency Division Multiple Access
FBER	Frame Bit Error Rate
FHSS	Frequency Hopping Spread Spectrum
FIFO	First Input First Output
FIR	Finite Impulse Response
FM	Frequency Modulation
FMIPv6	Fast Handovers for Mobile IPv6
FN	Foreign Network
FQDN	Fully Qualified Domain Name
FR	Full Rate
FRAMES	Future Radio Wideband Multiple Access System
FSK	Frequency Shift Keying
FSN	Fragment Sequence Number
FTChs	Forward Traffic Channels
FTP	File Transfer Protocol
GIF	GPRS Interworking Function
GPRS	General Packet Radio Services
GGSN	Gateway to GPRS Support Node
GMH	Generic MAC Header
GMM	GPRS Mobility Management
GMSK	Gaussian Minimum Shift Keying
GMSC	Gateway Mobile Switching Center
GRE	Generic Routing Encapsulation
GRX	GPRS Roaming Exchange
GSM	Global System for Mobile Communication
GTP	GPRS Tunneling Protocol
GTT	Global Title Translation
HA	Home Agent
HAAA	Home AAA
HCF	Hybrid Coordination Function
HCA	Hybrid Channel Assignment
HFC	Header Error Check Field
HLR	Home Location Register

HMIPv6	Hierarchical Mobile Internet Protocol version 6
HN	Home Network
HoA	Home Agent Address
HR	Half Rate
HSS	Home Subscriber Server
HSCSD	High Speed Circuit Switched Data
HTTP	Hiper Text Transfer Protocol
IAPP	Inter Access Point Protocol
IBSS	Independent Basic Service Set
LCIA	Loosely Coupled Integration Architecture
ICMP	Internet Control Message Protocol
ICMPv4	Internet Control Message Protocol version 4
ICMPv6	Internet Control Message Protocol version 6
ID	Identifier
IDRP	ICMP Router-Discovery Protocol
IEEE	Institute of Electrical and Electronics Engineering
IETF	Internet Engineering Task Force
IFS	Interframe Spacing
IIR	Infinite Impulse Response
IMS	IP Multimedia Subsystem
IMSI	International Subscriber Identity Module
IMEI	International Mobile Equipment Identity
IMT	International Mobile Telephone
IP	Internet Protocol
IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
IP-M IP	Multicast
IR	Infra Red
IRC	Internet Roaming Client
ISDN	Integrated Services Digital Networks
ISI	Inter Symbol Interference
ISN	Initial Sequential Number
ISUP	ISDN user Part
ISM	Industrial, Scientific and Medical frequency band
ISP	Internet Service Provider
IS-95	Interim Standard 95
IS-48	Interim Standard 48
ITAR	International Traffic and Arms Regulation
ITU	International Telecommunication Union
ITU-DS	ITU Direct Sequence
ITU-FT	ITU Frequency Time
ITU-MC	ITU Multi Carrier
ITU-SC	ITU Single Carrier
ITU-TC	ITU Time Code
IWF	Interworking Function
IWU	Interworking Unit
L2TP	Layer-2 Tunneling Protocol

LA	Location Area
LAC	Location Area Code
LAI	Location Area Identifier
LAN	Local Area Network
LAPD	Link Access Protocol D
LAPDm	Link Access Protocol D in ISDN
LCoA	On-Link CoA
LCIA	Loose Coupling Integration Architecture
LED	Light Emitting Diode
LIG	Lawful Interception Gateway
LINA	Location Independent Network Architecture
LIN6	Location Independent Network Architecture for IP version 6
LLC	Logical Link Control
LSS	Location Services SupportLOS Line of Sight
LPD	Link Protocol Discriminator
MAC	Medium Access Control
MAI	Multiple Access Interference
MAP	Message Application Part
MAP	Mobility Anchor Point
MAP	Maximum a Posteriori Probability
MAST	Multiple Address Service for Transport
MC-CDMA	Muticarrier CDMA
MCC	Mobile Country Code
ME	Mobile Equipment
MFSK	Minimum Frequency Shift Keying
MGCP	Media Gateway Control Protocol
MGW	Media Gateway
MGCF	Media Gateway Control Function
MIN	Mobile nodes Identification Number
MIP	Mobile Internet Protocol
MIPv4	Mobile Internet Protocol version 4
MIPv6	Mobile Internet Protocol version 6
MLME	MAC Layer Management Entity
MLS	Maximal Length Sequence
MLSE	Maximum Likelihood Sequence Equalizer
MM	Mobility Management
MMSE	Minimum Mean Square Error Equalizer
MN	Mobile Node
MNC	Mobile Network Code
MPDU	MAC Protocol Data Unit
MPSK	M-ary Phase Shift Keying
MS	Mobile Subscriber
MSC	Mobile Switching Center
MSDU	MAC Service Data Units
MSK	Minimum Shift Keying
MSISDN	Mobile Station ISDN Number
MSRN	Mobile Station Roaming Number
MT	Mobile Terminal

MTP	Message Transport Protocol
NABP	Node B Application Part
NAI	Network Access Identifier
NAS	Non Access Stratum
NAV	Network Allocation Vector
NIC	Network Interface Card
NFS	Network File System
NAAP	Network Access Authentication and Accounting Protocol
NSAPI	Network Service Access Point Identifier
NLOS	Non-Line of Sight
NLUM	Neighbour List Update Message
NMT	Nordic Mobile Telephones
N-PDU	Network Layer Packet Data Unit
NSS	Network and Switching Subsystem
NTT	Nippon Telephone and Telegraph
OCS	Online Charging System
OLPC	Open Loop Power Control
OMC	Operating and Maintenance Centre
OSA	Open System Authentication
OSI	Open System Interconnection
OSS	Operation Subsystem
OSPF	Open Shortest Path First
OFDM	Orthogonal Frequency Division Multiplex
OWLAN	Operator's Wireless Local Area Network
PACS	Personal Access Communication Systems
PAN	Personal Area Network
PAS	Personal Access System
PACCH	Packet Associated Control Channel
PAGCH	Packet Access Grant Channel
PBCCH	Packet Broadcast Control Channel
PBX	Private Branch Exchange
PCF	Point Coordination Function
PCH	Paging Channel
PCM	Pulse Code Modulation
PCU	Packet Control Unit
PCPCH	Physical Common Packet Channel
PDA	Personal Digital Assistance
PDC	Personal Digital Cellular
PDF	Probability Density Function
PDF	Power Delay Profile
PDN	Packet Data Network
PDP	Packet Data Protocol
PDCP	Packet Data Convergence Protocol
PDTCH	Packet Data Traffic Channel
PDU	Protocol Data Unit
PDGW	Packet Data Gateway

PER	Provider Edge Router
PG	Processing Gain
PHS	Personal Handyphone System
PhSF	Payload Header Suppression Field
PHY	Physical Layers
PIFS	Point Coordination Function (PCF) Inter Frame Spacing
PL	Path Loss
PLCP	Physical Layer Convergence Protocol
PLME	PHY Layer Management Entity
PLMN	Public Land Mobile Telephone Networks
PMD	Physical Medium Dependent
PMM	Packet Mobility Management
PNCH	Packet Notification Channel
PN	Pseudorandom Noise
PO	Portal
PPCH	Packet Paging Channel
PPDU	Packet Protocol Data Unit
PPP	Point-to-Point Protocol
PRACH	Packet Random Access Channel
P-CCPCH	Primary Common Control Physical Channel
PTCCH	Packet Traffic Control Channel
PSTN	Public Switched Telephone Network
PS	Packet Switching
PSK	Phase Shift Keying
PSMM	Pilot Strength Measurement Message
PTP	Point to Point
PTM	Point to Multipoint
PTM-M	PTM-Multicast
P-TMSI	Packet Temporary Mobile Subscriber Identity
QAP	QoS Enhanced Access Point
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QSTA	QoS Enhanced Stations
QPSK	Quadrature Phase Shift Keying
QPCh	Quick Paging Channels
RADIUS	Remote Authentication Dial-In-User
RAN	Radio Access Network
RF	Radio Frequency
RLC	Radio Link Control
RA	Routing Area
RAND	Random Number
RAB	Radio Access Bearer
RACh	Reverse Access Channel (RACh)
RAI	Routing Area Identifier
RAN	Radio Access Network
RANAP	Radio Access Network Application Part
RARP	Reverse Address Resolution Protocol

RACCH	Random Access Control Channel
RACH	Random Access Channel
RCoA	Regional CoA
RFC	Request For Comment
R4	Release 4
R99	Release 99
RLC	Radio Link Control
RIP	Routing Information Protocol
RNC	Radio Network Control
RR	Radio Resource
RRC	Radio Resource Control
RRM	Radio Resource Management
RNSAP	Radio Network Subsystem Application Part
RSS	Radio Subsystem
RSVP	Resource Reservation Protocol
R-SGW	Routing Signaling Gateway
RTCh	Reverse Traffic Channel
RTS	Request To Send
RTT	Radio Transmission Technology
SACCH	Slow Associated Dedicated Control Channel
SACH	Slow Associated Control Channel
SAP	Service Access Point
SAPI	Service Access Point Identifier
SBS	Symbol-to-Symbol
SCCP	Signaling Connection Control Part
SCP	Service Control Point
SCTP	Stream Control Transmission Protocol
SCH	Synchronisation Channel
SCCPCH	Secondary Common Control Physical Channel
SDCCH	Standalone Dedicated Control Channel
SDMA	Space Division Multiple Access
SDR	Software Defined Radio
SE	Sequence Estimator
SFD	SYNC Field Delimiter
SGN	Service Gateway Network
SGSN	Serving GPRS Support Node
SMG	Secure Mobility Gateway
SIFS	Short Inter Frame Spacing
SLA ID	Service Level Agreement Identifier
SNMP	Simple Network Management Protocol
SINAD	Signal-to Noise and Distortion Ratio
SIP	Session Initiation Protocol
SISN	Source Initial Sequential Number
SIR	Signal-to-Interference Ratio
SLR	SGSN Location Register
SIM	Subscriber Identity Module
SLA	Service Level Agreement

SM	Session Management
SMS	Short Message Services
S/N	Signal-to-Noise ratio
SNDCP	Sub-Network Dependent Converged Protocol
SNMP	Simple Network Management Protocol
SRES	Signed Response
SRNS	Serving Radio Network System
SS7	Signaling System Number 7
SSP	Service Switching Point
SSDT	Site Selection Diversity Transmission
STP	Service Transfer Point
STDMA	Synchronous Time Division Multiple Access
SyncCh	Synchronisation Channel
SRNC	Serving Radio Network Control
SUAL	SS7 SCCP User Adaptation Layer
TACS	Total Access Communication System
TCH	Traffic Channel
TCAP	Transaction Capabilities Application Part
TCP/IP	Transport Control Protocol/Internet Protocol
TCIA	Tightly Coupled Integration Architecture
TFCI	Transport Format Combination Indicator
TDD	Time Division Duplex
TDM	Time Division Multiplex
TDMA	Time Division Multiple Access
TIA	Telecommunication Industry Association
TID	Tunnel Identifier
TE	Terminal Equipment
TEID	Tunnel End Identifier
TFI	Traffic Flow Identifier
TLDN	Temporary Local Directory Number
TLLI	Temporary Logical Link Identifier
TMSI	Temporary Mobile Subscriber Identity
TPC	Transmit Power Control
TRAU	Trans-coding and Rate Adaptation Unit
TS	Time Slot
T-SGW	Transporting Signaling Gateway
TRX	Transceiver
TUP	Telephone User Part
Tx-Rx	Transmitter-Receiver
UL-MAP	Uplink MAP
UARFCN	UTRA Absolute Radio Frequency Channel Number
UAF	User Adaptation Function
UCD	Uplink Channel Descriptor
UDP	User Datagram Protocol
UE	User Equipment
UIUC	Uplink Interval Usage Code
UMTS	Universal Mobile Telecommunication Systems
UNI	User to Network Interface

NNI	Network to Network Interface
URA	UTRAN Routing Area
USIM	UTRAN Subscriber Identity Module
UTRAN	Universal Terrestrial Radio Access Network
VLR	Visiting Location Register
VPN	Virtual Private Network
VoIP	Voice Over IP
VSA	Virtual Single Account
WAG	Wireless Access Gateway
WAN	Wide Area Network
WAP	Wireless Application Protocol
WAPN	WLAN Access Point Network
WCO	Weighted Carrier Ordering
WiMAX	Worldwide Inter operability for Microwave Access
WCDMA	Wideband Code Division Multiple Access
WEP	Wired Equivalent Privacy
WLAN	Wireless Local Area Network
WLAN AG	WLAN Access Gateway
WG	Working Group
WMM	WLAN Mobility Management
WSSUS	Wide Sense Stationary Uncorrelated Scattering
WTDMA	Wideband Time Division Multiple Access
WPAN	Wireless Personal Area Network
WWW	World Wide Web
ZF	Zero Forcing
XRES	EXpected user RESponse

Index

Symbol

A

B

C

D

E

F

G

H

I

K

L

M

N

O

P

Q

R

S

T

U

V

W

Z