Mobile Computing

Sipra DasBit
Professor
Department of Computer Science and Technology
Bengal Engineering and Science University Shibpur

Biplab K. Sikdar
Assistant Professor
Department of Computer Science and Technology
Bengal Engineering and Science University Shibpur

PHI Learning Private Limited
New Delhi-110001
2009

CONTENTS

PREFACE

Recently, in the curricula of several universities, Mobile Computing is included as a core/elective paper in B.E./B.Tech. in Computer Science and Engineering and allied disciplines, MCA and B.Sc (computer science). Thus students from different backgrounds have to study this subject approaching from different levels. The available textbooks on Mobile Computing are organized to suit best the needs of the students of Electronics and Tele-communication Engineering. The topics covered are biased to Electronics courses, and put more stress on characterization of wireless channels, detailing the various modulation techniques, speech coding, etc. A sound grounding in Digital Communication is the prerequisite for such texts. Therefore, the contents are not so easy to follow for the students of other engineering branches. Further, even the books that are not so biased to Communication Engineering target the telecom professionals and cover mainly the commercial aspects, services, applications, and standards rather than concentrate on the theoretical concepts. As teachers, we find it difficult to refer only one textbook on Mobile Computing to our students.

This book is aimed to support the students who lack exhaustive knowledge in Communication Engineering but have opted for Mobile Computing as a subject. It is planned in such a manner that the concept of Mobile Computing can be gleaned even without prerequisite knowledge of Communication Technology. In a conscious effort to keep the book handy and free from technical jargons, we have chosen only one implementation standard from each category of Wireless Networks, and developed the contents based on that.

Chapter 1 introduces the evolution of mobile communication services starting from wireless LAN to 3G services. It then includes the fundamentals

1

INTRODUCTION

The seed of today's mobile computing was implanted long back in 1897 since G. Marconi invented the radio's ability to provide continuous contact with ships in the English Channel. The seed was grown up to a sapling and finally turned into a huge tree. This is due to the improvements in digital and RF circuit fabrication, a large-scale circuit integration and other miniaturization technologies. The mobile computing spreads its branches providing high speed text and multimedia data services as well as extends its capacity in terms of subscriber accommodation. It enables today's cheap, improved, portable and enriched services available for the entire population. This reality is of many years' effort in different wings of technology.

1.1 HISTORY

Prior to the analog cellular system was in operation (Chicago, 1983), the mobile radio systems were based on trunking principle. Designers provided optimal number of channels for sharing among the users. The available frequency spectrum (150–450 MHz) was divided into a number of frequency channels. The channels were shared by the users. Such a system used single high-powered transmitter installed on a large tower. The transmitter was powerful enough to cover a distance of 50 km. However, the concerns of those systems were:

- the mobile receivers (generally installed in automobile) were of large size and therefore, difficult to carry.
- a very limited number of channels were used. Even for call termination, a mobile station had to compete for a channel.

- the call blocking probability was high.
- the coverage of a network was limited.
- the number of simultaneous calls in progress was very restricted.

In spite of radio's (wireless) tremendous ability to communicate anytime anywhere, the concept of wireless communication was not so popular in a mass scale. During the 1960s Bell Laboratories developed the cellular concept. A major breakthrough took place with the introduction of frequency reuse mechanism in cellular operation. Simultaneously, the advancement in IC technology, digital signal processing and battery technology aided the sharp growth in mobile and personal communication services. Such systems were primarily analog and were best known as *1st generation mobile systems*.

Around late 1980s the Wireless Local Area Network (WLAN) technology, an alternative to wired LAN, was launched. The two prominent standards for WLAN emerged were the IEEE 802.11 in the US and the High Performance European Radio LAN (HIPERLAN) in Europe. In 1990, the first digital cellular specification was released for the Global System for Mobile (GSM) communication and was implemented in Europe. It operates in 935–960 MHz and in 890–915 MHz bands. The systems using digital cellular technology are perceived as the *2nd generation mobile* or Personal Communication Services (PCS) systems. The notable among the 2nd generation systems are the Digital European Cordless Telephone (DECT) and Digital Communication Systems (DCS, 1800 MHz version of GSM) introduced in Europe, IS-95 in North America, and Personal Digital Cellular (PDC) in Japan.

The initial focus of the 2nd generation mobile systems was on circuit-switched voice and low bit rate data services. In circuit-switched connection, a dedicated connection is established for the entire duration of a call. In cellular network, circuit-switched voice demands a dedicated radio channel between a handset (mobile station) and its nearest network tower (base station) and a dedicated phone line between the terminal point of the network (mobile switching centre) and telephone network (public switch telephone network), even when the mobile station moves to a different place under the new network tower.

Over the years, the demand to have higher data rates increases for applications such as Internet access, various standards like General Packet Radio Service (GPRS) for GSM, Cellular Digital Packet Data (CDPD), High Speed Circuit-Switched Data (HSCSD), etc. These are considered as gateway to next generation wireless access and, therefore, referred to as *2.5 generation standards/systems*. However, the 2nd and 2.5 generation systems are lacking in

- Interoperability among cellular, cordless, satellite, LANs, etc.
- Handling multimedia services along with voice, data and Internet.
- Global mobility—that is, global seamless roaming.
- Quality of Service (QoS).

Although the first rush to wireless was for voice transmission, attention was also paid to data services demanding the wireless Internet. Two technologies/standards, Mobile IP and Wireless Application Protocol (WAP), were proposed to provide application level support for wireless networking.

The Mobile IP was conceived around 1992 and standardized in 1996. It was targeted to enable seamless roaming for Internet connected devices. Simultaneously, around 1997 the WAP was introduced and then implemented in 2000. With the WAP a wireless service provider could provide Internet services to mobile devices over the same voice network. The evolution of mobile communication services is recorded in Figure 1.1.

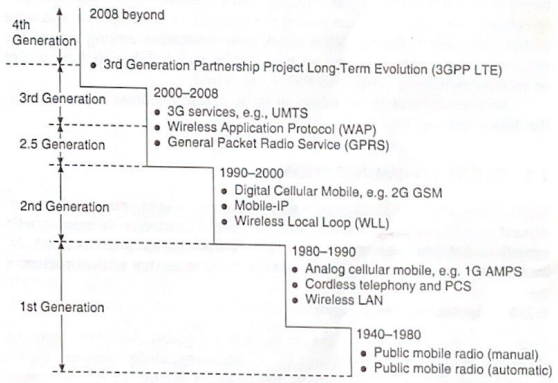


Figure 1.1 Evolution of mobile communication services.

To meet the challenges of 2nd generation mobile systems, the International Telecommunication Union (ITU) has developed a standard IMT-2000 and initiated the attempt towards development of the 3rd generation mobile systems. The 3rd generation systems accommodate the existing wireless access systems (that are the harvest of many efforts from different sectors) under one umbrella. The Universal Mobile Telecommunication Systems (UMTS), proposed by European Telecommunications Standard Institute (ETSI), is developed at par with the IMT-2000. It supports global roaming facility as well as the high-speed data services for different QoS traffic classes ranging from the ordinary background data (e.g. e-mail) to the most essential conversational data (e.g. voice) or even the highly prioritized multimedia services (e.g. video on demand).

The higher data rate is projected from *4th generation mobile systems*. It will be able to provide an all IP network where voice, data and streamed

multimedia can be given to users on an anytime anywhere basis, and at higher data rates than 3rd generation. The international telecommunications regulatory and standardization bodies are working for commercial availability of 4G networks around 2012–2015. One of the technologies which are being considered as pre-4G is 3rd Generation Partnership Project Long Term Evolution (3GPP LTE). The LTE project is not a standard, but it will result in the extensions and modifications of the UMTS system. It is to improve the UMTS mobile phone standard to cope up with future technology evolutions.

In all mobile communication networks, at least one hop (if not all hops) of communication is the wireless radio communication. For example, in cellular network, communication between the mobile station and base station is wireless. On the other hand, communication among rest of the network components, e.g. between base station and mobile switching centre or mobile switching centres to PSTN are wired.

An introduction to the basics of radio communication is provided in the following section.

1.2 RADIO COMMUNICATION

Radio communication is a popular way of transmitting information between distant places. The unit of wireless radio communication is electromagnetic signals and it provides means to exchange information such as symbols, intelligence between two points—that is, from a source to destination.

1.2.1 Electromagnetic Signal

Light is a form of electromagnetic wave/signal. Another form of electromagnetic signal is used in radio broadcasting. Electromagnetic signal is oscillatory in nature. The graph displayed in Figure 1.2 is a snapshot representing the strength of a signal. The distance between two wave crests is the wavelength of this signal. The signal traverses at the velocity of light.

A signal is qualified by its frequency of oscillation—that is, the number of oscillations per second. The usual frequency we get from the electric currents in our household connection is about 50/60 cycles per second. In radio (AM) broadcasting, frequency of the electromagnetic signal used is in the order of 10^6 cycles per second. The frequency used for FM radio/TV is more than the radio broadcast and short waves/radar frequency is more than the frequency used for FM radio/TV. For reference, it can be stated that the further increase in the frequency of oscillation make electromagnetic signal visible by human eyes. The visible frequency is in the range of 5×10^{14} to 5×10^{15} cycles per second.

The unit of frequency measure is Hertz (Hz); 1Hz represents the frequency of one cycle per second. The megahertz (MHz) is the 10^6 Hz and gigahertz (GHz) implies 10^9 Hz. That is, 1 MHz indicates that the signal completes one million (10^6) cycles per second and 1 GHz implies 10^9 cycles per second.

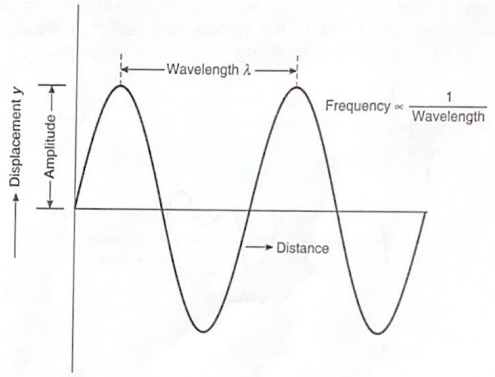


Figure 1.2 A typical electromagnetic signal.

The variable quality of the electromagnetic waves could also be the wavelength. The frequency is inversely proportional to the wavelength. That is, higher frequencies have the shorter wavelengths, while lower frequencies indicate larger wavelength. The frequency (f) can be represented as $f = v/\lambda$, where v = phase velocity of wave and λ = wavelength. In vacuum, v is the velocity of light. The radio waves can be of wide range of wavelengths. The regular radio broadcasting wavelength is about 500 metres. The wavelength of short-wave/radio wave is in the order of millimetre. Further, the frequency $f = 1/T$, where T (period) is the time between two consecutive occurrences of the event in wave propagation. For example, if period T of a signal is 1s (10^0 s), then the frequency $f = 1$ Hz (10^0), and if $T = 1$ ms (10^{-3} s), $f = 1$ kHz (10^3 Hz).

Analog vs digital signal

The signal can be either analog or digital. An analog signal is one in which the signal intensity varies in a smooth fashion (continuous) over time. On the contrary, a digital signal is one in which the signal intensity maintains a constant level (discrete) for some period of time and then changes to another constant level. Figure 1.3 shows a sample of analog and digital signals.

[1] The phase velocity of a wave is the rate at which the phase of the wave propagates in space.

Spectrum/bandwidth

The electromagnetic waves cover a wide frequency range. The spectrum of a signal is the range of frequencies that it contains. If a frequency spectrum ranges from f to $3f$, the absolute bandwidth of the signal is $3f - f = 2f$, the width of the spectrum.

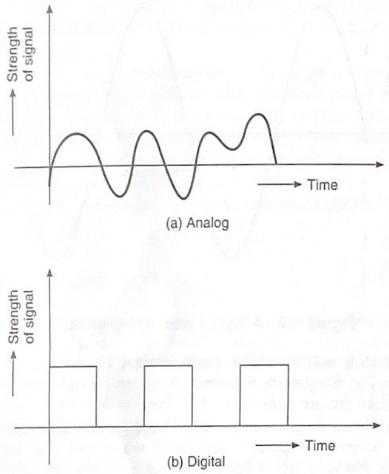


Figure 1.3 Signal.

Many signals have an infinite bandwidth. However, most of the energy contains within a relatively narrow band of frequencies. This narrow band is referred to as the *effective bandwidth* or simply *bandwidth*. In radio, the width of frequency spectrum can impart more information in communication process.

Characteristics of electromagnetic signals

Table 1.1 shows the spectrum of electromagnetic signals with common nomenclature. The signals have different characteristics and abilities for communication. The key considerations are: how far the waves can go, through what these can go through and how much data these can carry. All these have an impact on how useful a signal will be for different types of wireless communication. In the current context, the main concern is on radio portion (a few cycles to 10^9 cycles per second) of the spectrum (Refer to Figure 1.4) that is used in mobile computing.

Table 1.1 Radio Frequency/Bandwidth

Frequency	*Band*	*Type of radiation*
< 30 kHz	Very Low Frequency (VLF)	Long radio wave
30–300 kHz	Low Frequency (LF)	Long radio wave
300 kHz–3 MHz	Medium Frequency (MF)	Long radio wave
3–30 MHz	High Frequency (HF)	Short radio wave
30–300 MHz	Very High Frequency (VHF)	Short radio wave
300 MHz–3 GHz	Ultra High Frequency (UHF)	Short radio wave
3–30 GHz	Super High Frequency (SHF)	Microwave
> 30 GHz	Extremely High Frequency (EHF)	Infrared ... Gamma rays

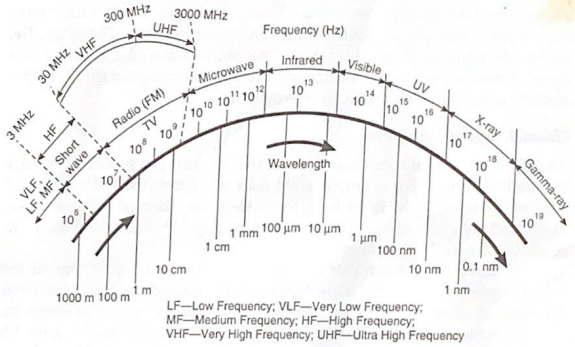


Figure 1.4 The electromagnetic spectrum.

Range of transmission

High frequencies (HF) are reflected by the ionosphere and bounce back and forth between the ionosphere and the earth's surface. As a result the HF signal can convey information/data to thousands of kilometres. Further, a HF signal can penetrate obstructions such as buildings, mountains, etc. Such a frequency is typically used by the radio stations to broadcast programs worldwide.

The higher frequencies such as SHF and EHF normally operate on a line-of-sight (LOS) basis. In LOS, the receiver receives the signal from the transmitter on a straight line. These frequencies do not bend around the earth's surface like HFs and, therefore, require repeaters, every few tens of kilometres, to establish links between the nodes.

Microwaves (SHFs) can pass right through some objects but are blocked by relatively solid objects like buildings and metal structures. The light waves behave like microwaves but light waves are even more

restricted in distance travelled and comparatively less powerful to penetrate objects.

The range of transmission depends not only on the characteristics of a signal but also on the parameters such as power of the transmitter, sensitivity of receiver, antenna types, antenna height, quality of the transmission medium used, method used to send information, amount of interference present, atmospheric conditions, etc.

Information-carrying capacity

There is a direct relationship between the information-carrying capacity of a signal and its bandwidth. The information-related modifications made to the signal are normally tied to its frequency. The range of frequencies a signal contains is the bandwidth. That is, a larger bandwidth means more information can be sent for a given period of time. Therefore, the greater the bandwidth, the higher the information-carrying capacity of a signal. For example, in broadband, information-carrying capacity (bits per second) is far more than that in narrowband.

Channel capacity and noise

An individual communication path that carries signals at a specific frequency is called a channel. For example, GSM uses two bands (890–915 MHz and 935–960 MHz) of 25 MHz each for the system use. Each of 25 MHz bands is divided into 200 kHz wide channels resulting 125 channels. That is, in total, GSM uses 250 channels.

A variety of impairments can distort or corrupt a signal. One of the major impairments is due to noise. Noise is the unwanted signal that combines to a desired signal and distorts it during transmission as well as at reception.

A noise affects and limits the data rate in transmission. Figure 1.5 shows how a transmitted signal is affected by noise in transmission medium. The 'channel capacity' qualifies such effect of noise. The maximum rate at which the data can be transmitted over a channel, under given conditions, is referred to as the 'channel capacity'. It is the upper bound on the amount of information that can be reliably transmitted over a communication channel. On the other hand, 'signal strength' is the magnitude of electric field at a reference location around the transmitting antenna.

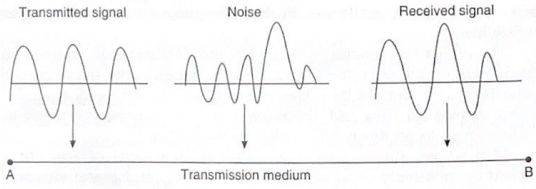


Figure 1.5 Signal distortion due to noise.

The greater signal strength improves the ability to receive data correctly in the presence of noise. The signal-to-noise ratio (SNR), the ratio of power in a signal to the noise power (corrupting the signal at a particular point of transmission), is the measure of data received correctly and is defined as

$$\text{SNR} = \left(\frac{P_{\text{signal}}}{P_{\text{noise}}}\right) = \left(\frac{A_{\text{signal}}}{A_{\text{noise}}}\right)^2$$

P refers average power and A is the root mean square amplitude.

It is usually expressed in terms of logarithmic scale. In decibel it is 10 times the logarithm of power ratio. That is,

$$\text{SNR}_{\text{dB}} = 10 \log_{10}\left(\frac{\text{Signal power}}{\text{Noise power}}\right)$$

For example, in a cellular system, if the signal power/noise power = 96.2, the SNR ≈ 20 dB. A high value of SNR means signal received is of high quality. The SNR is also referred to as the signal-to-interference ratio (S/I).

1.2.2 Full-duplex Radio Communication

Based on the direction of communication, a radio communication system can be referred to as either simplex or duplex. A simplex system allows

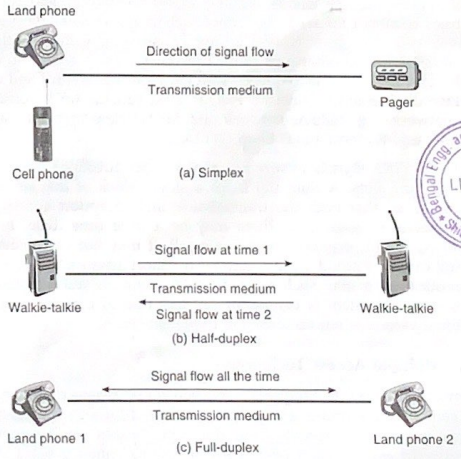


Figure 1.6 Radio communication.

communication in one direction. An example of simplex system can be the pager. A pager receives message but cannot reply/acknowledge. A duplex system, on the other hand, allows two-way communication between two points. In half-duplex radio system, the two-way communication cannot occur simultaneously. The same half-duplex channel can be used for transmission as well as reception; but at any point of time either transmission or reception can take place. Walkie-talkie is an example of half-duplex system. Whereas, the full duplexing is a mechanism by means of which simultaneous transmission and reception is possible in a communication system. In other words, communication is possible in both directions simultaneously. The conventional telephone system is the right example of full-duplex system, where simultaneous exchange of dialogue is possible. Figure 1.6 shows various types of radio communication.

In wireless systems, full duplexing is implemented either by providing two separate channels (one for transmission and the other for reception) or by providing two adjacent time slots on a single radio channel. The former method for achieving full-duplex communication is called the Frequency Division Duplexing (FDD) and the latter one is called the Time Division Duplexing (TDD).

FDD: In case of FDD, signals are differentiated based on the frequencies. Two distinct set of frequencies are assigned for this purpose. Each such set of frequencies corresponds to a simplex channel in FDD and, therefore, a duplex channel consists of two simplex channels. In a cellular network, a device called duplexer is used in mobile station as well as in the base station allowing simultaneous radio transmission and reception on the duplex channel pair. The frequency split between the forward and reverse (backward) channels is constant. FDD is most suitable for wireless wide area network, e.g. cellular network and for wireless metropolitan area network, e.g. Wireless Local Loop (WLL).

TDD: In TDD, signals in forward and reverse directions are assigned separate time slots. A time slot is for a single block of frequencies and kept small so that both the transmission and reception appear to be simultaneous to a user. As there may be a little time delay between transmission and reception, however small, it may not be considered a true full-duplex method. TDD is suitable for short distance and low power communication system such as cordless telecommunication system. The cost of such a system is comparatively less due to a fewer wires and simplified receivers required for the implementation.

1.2.3 Multiple Access Techniques

The 'multiple access' technique is a mechanism by means of which many users can share the frequency spectrum allocated for an application. The systems specify how signals from different sources can be combined efficiently for transmission over a given radio frequency band. At the receiving end, the signals are then separated out ensuring minimum mutual

interference. There are mainly three multiple access techniques—Frequency Division Multiple Access (FDMA), Time Division Multiple Access (TDMA), and Code Division Multiple Access (CDMA). FDMA system separates the total available bandwidth into several non-overlapping smaller bands/channels. Each channel has the ability to support a user. TDMA system splits users into an available pair of channels, but it also assigns each user an available time-slot within that channel. In each slot, only one user is allowed to either transmit or receive. With CDMA, unique digital codes, rather than the separate radio frequencies/channels, are used to differentiate users. A specific code is assigned to each user and only that code can demodulate the transmitted signal. Each user utilises the entire block of allocated spectrum space to carry the message. An elaborate discussion on these techniques are provided in Chapter 3 as the part of digital communication.

1.2.4 Operational Modules for Radio Communication

In this subsection, the major tasks involved in radio communication are identified. Figure 1.7 shows the operational modules in both the transmitter and receiver ends. When a signal, i.e. voice, video, etc. is generated for communication, it is processed in several phases prior to transmission. At first the analog signal is transformed into digital by the source coder. The channel coder detects and/or corrects the errors due to transformation. However, all the errors cannot be corrected by the channel coder. The effect of such errors is minimized by the interleaver. Finally, the signal is given upshift of frequency so that it can be transmitted over the radio path in an efficient manner. When the signal is received at the receiver, a reverse process is applied to extract the original signal.

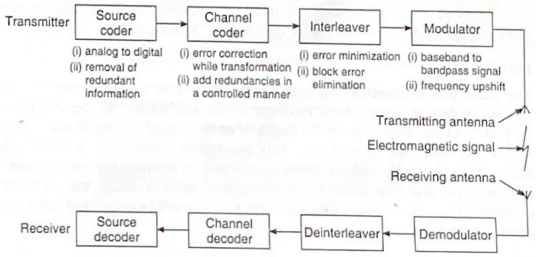


Figure 1.7 Operational modules for radio system.

Source coding

In this phase the input signal to be transmitted (voice, video, etc.) is transformed from analog to digital. In wireless communication, there exists

a scarcity of bandwidth. However, for applications like voice communication, high quality service is needed. Therefore, the challenge in such speech coding system is to ensure high quality speech transmission exploiting the least possible channel capacity. It demands a trade-off between these two conflicting requirements. In order to maximize spectral efficiency[2], the more efficient voice coding methods are introduced in cellular systems. Once analog signal is converted to digital, the encoded data bits, called source bits, carry information. Due to a typical coding mechanism, some of the source bits have more impact than others and therefore need to be protected from errors. On the other hand, to save the bandwidth, bits carrying redundant information should be removed prior to transmission. So, the removal of redundant information is also the task of a source coder.

Channel coding

After removal of redundancies by the source coder, in a radio system, the source bits pass through the channel coding process. It improves quality of transmission by protecting source data from errors caused by the events such as multipath fading (transmitted signals take multiple paths to reach to the receiver causing interference between multiple versions of the transmitted signal having differences in amplitude and phase received at slightly different times), Doppler shift (due to movements), etc. The data/information is protected from errors through introduction of redundancies in a controlled manner. However, the errors such as block error caused by fading cannot be efficiently corrected by the channel coding schemes and demand interleaving. A block error results in alteration of a sequence of bits (block of bits) from the transmitted bits. If the alteration can be spread within the whole bit stream, chances of loss of message are reduced. Interleaving does the task of spreading the loss of bits. Some of the channel coding techniques used in mobile communication are convolution coding, parity check coding and block coding.

Interleaving

The main task of interleaving is to protect the signal against block errors. In case of fading, errors tend to occur in blocks. Normally, the length of a block of errors depends on the depth of encountered fading. It scrambles and/or spreads the source data bits essentially randomizing the block errors. However, this process introduces delay in transmitted information. For applications like voice communication, such a delay may not be acceptable. So, the system designer has to trade-off between the removal of errors and delay.

Modulation

It is necessary that the source data needs to be transformed in a manner suitable for transmission in wireless communication system. Wireless

[2] Spectral efficiency is the amount of information that can be transmitted over a given bandwidth.

medium allows only analog transmission. If digital source data has to be transmitted over such a medium, the modulation is required. The modulation translates a source or baseband message signal to a bandpass signal (whose frequency content is concentrated in a narrow-band) at frequencies very high compared to a baseband frequency. The source or baseband message signal is called modulating signal and the generated bandpass signal is the modulated signal. The high frequency signal that used to translate the baseband message signal to the bandpass signal is called the carrier signal (Figure 1.8(a)).

There are various reasons for which the baseband signal cannot be directly transmitted in a wireless medium. The important one is the antenna height. It must be in order of magnitude of the signal's wavelength that is to be transmitted. For example, even for 1 MHz signal, antenna height must be in order of hundred metres or so. Therefore, the modulation phase basically performs two tasks—(i) digital modulation, i.e. translation of source encoded and channel encoded digital data into square wave binary pulses. This is the baseband signal if presented in frequency domain, (ii) frequency shifting. Because of signal attenuation, impractical antenna height, etc., it is necessary to shift the baseband signal spectrum to reside at a much higher centre frequency of the baseband signal. The three basic methods for the digital modulation are amplitude shift keying (ASK), frequency shift keying (FSK) and phase shift keying (PSK). Similarly, there are three analog modulation techniques—amplitude modulation (AM), frequency modulation (FM) and phase modulation (PM). The digital version of amplitude modulation (AM), frequency modulation (FM) and phase modulation (PM) are the amplitude shift keying (ASK), frequency shift keying (FSK) and phase shift keying (PSK) respectively.

AM: Amplitude modulation alters the amplitude of a carrier signal so that its amplitude matches with the original, i.e. baseband modulating signal (Figure 1.8(a)). That is, carrier is modulated such that its amplitude varies with the amplitude of the baseband modulating signal.

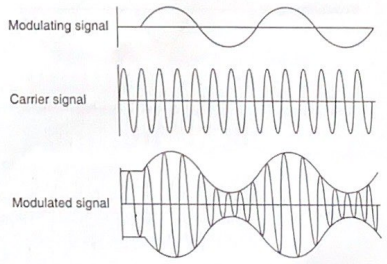


Figure 1.8(a) Amplitude modulation.

The frequency and phase of the carrier signal remain unaltered. It must be ensured that the frequency of the carrier must be much higher than the highest frequency in the baseband signal. The demodulation process at the receiving end filters out the carrier. The bandwidth required in modulated signal is twice of the bandwidth of a modulating signal.

AM works by varying the strength of the transmitted signal relating to the information being sent and suffers from power variation in the output.

ASK: The input to such a modulation process is the binary bit streams. The two binary values 1 and 0 are represented by two different amplitudes. Figure 1.8(b) shows the transformation of input bit stream 1001. In this example, the binary 0 is represented by amplitude 0. Here the carrier signal strength is varied in modulated signal. The frequency and phase of the carrier remains constant but amplitude is changed to match with the input binary signal(repeatation). However, various interferences in radio propagation such as multi-path propagation, noise, etc. affect amplitude greatly. So, ASK is normally not preferred in wireless radio transmission.

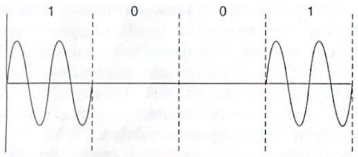


Figure 1.8(b) Amplitude shift keying.

FM: The frequency of the carrier signal is altered to carry the content of the modulating signal (Figure 1.9(a)). The other two parameters, i.e. amplitude and phase of the carrier are kept constant. As the career signal's

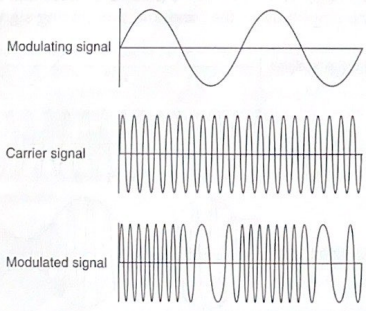


Figure 1.9(a) Frequency modulation.

amplitude remains unchanged, the FM is more immune to noise than AM and improves the overall signal-to-noise ratio (SNR) in the system. Unlike AM, the power output in FM is constant.

Frequency modulation is achieved by simply mixing the modulating and carrier signals. It follows the amplitude of the modulating signal. The range of frequency deviation in the modulated signal is proportional to the amplitude of modulating signal. For example, in modulation, if the carrier frequency is 1000 MHz and the modulating signal's amplitude is 20, then for 40 kHz frequency shift per unit amplitude, the modulated signal's frequency variation will be in between (1000 MHz – 20 × 40 kHz) and (1000 MHz + 20 × 40 kHz), i.e. 999.2 MHz and 1000.8 MHz.

The number of instances of such variation in frequency within the modulated signals are equal to the frequency of modulating signals. It means if the modulating signal frequency for the earlier example is 2 kHz, then the frequency variation (from 999.2 to 1000.8 MHz) in modulated signal will occur at the rate of 2000 times per second.

Further, the bandwidth for FM is 10 times of the bandwidth of modulating signal.

FSK: The simplest form of FSK is the binary FSK (BFSK). Instead of representing the input binary values by two different amplitudes, the BFSK represents 1 and 0 by the two different frequencies that are close to the carrier frequency. The FSK representation of example bit stream 1001 is shown in Figure 1.9(b). It can be noted that the peak amplitude and the phase remain constant. The FSK is more immune to noise than the ASK.

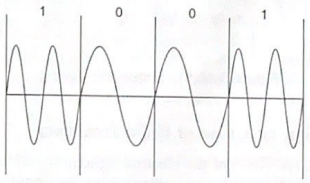


Figure 1.9(b) Frequency shift keying.

PM: In phase modulation, the carrier signal's amplitude and frequency remain constant after modulation. It is either advanced or retarded in its phase cycle (phase shift) by the modulating signal (Figure 1.10(a)) and the amount of phase shift is proportional to the amplitude of the modulating signal. The PM is less prone to the noise interference.

PSK: In phase shift keying, the phase of carrier signal is shifted in accordance with the change in input bit stream. Whenever there is a change in data, i.e. from 0 to 1 or 1 to 0, the phase is shifted by 180°. However, the amplitude and frequency remain constant. This is also a binary PSK (BPSK). The PSK

is more resistant to interference compared to the FSK. Figure 1.10(b) shows the PSK representation of the example bit stream 1001.

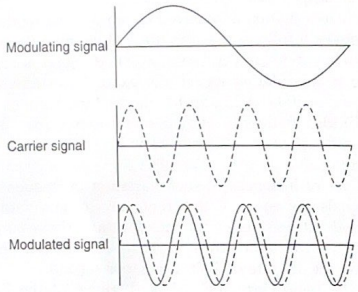


Figure 1.10(a) Phase modulation.

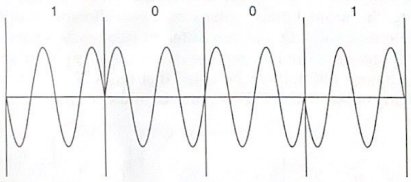


Figure 1.10(b) Phase shift keying.

1.2.5 Regulation on Usage of Radio Frequency

The spectrum (< 100 GHz) of the electromagnetic signal is prone to create interference with other services operating on the same frequencies. For example, short-range microwave transmissions operating on the same frequency must be kept separated by a reasonable distance. The avoidance of interference requires coordination between the users, and that eventually necessitates the regulatory boards or commissions.

There are agencies involved in setting the technical and maintenance standards, as well as laying down planning rules, performance targets and operational procedures for the management of telecommunication networks. Their target is to outline the recommendations of ITU (International Telecommunication Union) as the world's most authoritative body on telecommunications standards. One such organisation is International Frequency Registration Board (IFRB). It is the part of ITU

and is responsible for maintaining the list of radio frequencies used worldwide. It also conducts technical and planning studies for ITU.

In addition to the international bodies, some of the regional bodies such as FCC (Federal Communications Commission) and CEPT (Conference Europeene des Postes et des Telecommunication) are responsible for controlling specific ranges of the spectrum in the US and Europe respectively. The following are the issues involved in regulation of radio frequencies:

- Amount of power transmitted
- Physical configuration of the equipment and antenna
- Issuance of license to the user
- Annual renewal of license
- Setting up of an acceptable transmission system
- Maintaining compliance

1.3 SUMMARY

Summarily, if we compare with wired network, the counterparts in wireless can be made as wireless LANs, wireless MANs (metropolitan area network) and wireless WANs (wide area network). Wireless LANs (e.g. range—spanning two adjacent buildings) are categorized further as the personal area networks (Bluetooth, data rate 1Mbps, range 10 metre) and Business LANs (802.11b, data rate 11 Mbps, range 100 metre). Examples of wireless MANs (e.g. range—spanning a city) is the wireless Local Loop (data rate 37 Mbps to 1.25 Gbps). On the other hand, wireless WANs (e.g. range—spanning earth) having wide coverage can further be categorized as cellular networks (GSM, data rate 9.6 kbps), Satellite systems (Motorola Iridium, data rate 64 Mbps) and Paging networks (ReFLEX, data rate 6.4 kbps).

The book is intended to address different issues in wireless networks. As mobile computing covers computing issues related to the voice and data communication over wireless network, we have chosen one representative network from each category of wireless network and elaborated those throughout the book. To start with, a sizeable number of chapters discuss cellular network, an example network of the WAN category, including its implementation standards from 2 G to 2.5 G. The next few chapters are dedicated to describe example networks of the other categories, namely MAN and LAN along with their standards and applications. The rest of the chapters concentrate on mobile data services including the issues of WAP (Wireless Application Protocol) and Mobile IP. Finally, an overview of 3G cellular networks (e.g. UMTS) is introduced in the last chapter.

BIBLIOGRAPHY

Akaiwa Y., and Y. Nagata, "Highly efficient digital mobile communications with a linear modulation method", *IEEE Journal of Selected areas in communications (SAC)*, Vol. 5, No. 5, pp. 890–895, 1987.

Akeyama, A., T. Nagatsu and Y. Ebine, "Mobile Radio Propagation Characteristics and Radio Zone Design Method in Local Cities", *Review of the Electrical Communication Laboratories*, Vol. 30, No. 2, pp. 308–317, 1982.

Bertoni, H., *Radio Propagation for Modern Wireless Systems*, NJ, Prentice-Hall, 2000.

Bullington, K., "Radio Propagation Fundamentals", *Bell System Technical Journal*, NJ, Vol. 36, pp. 593–626, 1957.

Cox, D.C., "Wireless personal communications: what is it?", *IEEE Personal Communications*, USA, Vol. 2, No. 2, pp. 20–35, 1995.

Glover, I., and P. Grant, *Digital Communications*, NJ, Prentice-Hall, 1998.

Goodman, D.J., *Wireless Personal Communications Systems*, Addison-Wesley, Reading, MA, 1997.

Li, V.O.K., and X. Qiu, "Personal communication system (PCS)", *Proceedings of the IEEE*, Vol. 83, No. 9, pp. 1210–1243, 1995.

Padgett, J.E., C.G. Ginther, and T. Hattori, "Overview of wireless personal communications", *IEEE Communications Magazine*, Vol. 33, No. 1, pp. 28–41, 1995.

Proakis, J., *Digital Communications*, New York, McGraw-Hill, 2001.

Schiller, Jochen H., *Mobile Communications*, New Delhi, Pearson Education, 2007.

Sklar, B., *Digital Communications: Fundamentals and Applications*, NJ, Prentice-Hall, 2001.

Stallings, William, *Wireless Communications and Networks*, New Jersery, Pearson Education, 2006.

Steele, R. (Editor), *Mobile Radio Communications*, New York, IEEE Press, 1994.

Stuber, G.L., *Principles of Mobile Communications*, Boston, 2nd ed., Kluwer Academic Publishers, 2001.

Viterbi, A.J., and J.K. Omura, *Principles of Digital Communication and Coding*, McGraw-Hill College, 1979.

Wilson, S.G., *Digital Modulation and Coding*, Upper Saddle River, NJ, Prentice-Hall, 1996.

Xia, H., *et. al.*, "Radio Propagation Characteristics for Line-of-Sight Microcellular and Personal Communications", *IEEE Transactions on Antennas and Propagation*, Vol. 41, No. 10, pp. 1439–1447, 1993.

Xiong, F., *Digital Modulation Techniques*, Boston, Artech House, 2000.

Yacoub, M.D., *Foundations of Mobile Radio Engineering*, Boca Raton, FL, CRC Press, 1993.

Ziemer, R.E., and R.L. Peterson, *Digital Communications*, Prentice-Hall, Englewood Cliffs, 1990.

Ziemer, R.E., and R.L. Peterson, *Introduction to Digital Communications*, New York, Macmillan Publishing Company, 1992.

REVIEW QUESTIONS

1. Prepare a write-up describing evolution of mobile communication services.
2. Give an example of wireless system each from 1 G, 2 G, 2.5 G and 3 G.
3. Compare analog and digital signal.
4. Define wavelength and frequency of a signal. What is the relationship between these two parameters?
5. What is the relation between a signal's spectrum and its bandwidth?
6. How do you characterize an electromagnetic signal?
7. (a) What are the transmission ranges of the following signals?
   (i) High Frequency (HF) signal
   (ii) Super High Frequency (SHF) signal and Extremely High Frequency (EHF) signal
   (b) On which parameters transmission range of a signal depends?
8. What is information-carrying capacity of a signal? How is it related to bandwidth?
9. What is the concept of channel? Explain it with an example.
10. What is noise? How it affects signal transmission?
11. Define channel capacity and signal strength.
12. What is meant by SNR = 20 dB?
13. (a) Define simplex, half-duplex and full-duplex communication. Give examples of each such communication systems.
    (b) How full-duplexing is implemented in wireless system?
14. What is the significance of 'multiple access'? Name important multiple access techniques.
15. Draw the block diagram of a radio system at both the transmitter and receiver sides. Give a brief description of the tasks of each of the modules in the diagram.
16. What is the need of modulation? Differentiate between modulating and modulated signal.
17. What are the two major tasks of modulation? Explain.
18. What are the principles of AM and ASK, FM and FSK and PM and PSK?
19. How binary values are represented in frequency shift keying? Explain it citing an example.
20. What are the merits of FM over AM? In which way PM is advantageous over FM?
21. (a) Why regulation is essential on the usage of radio spectrum?
    (b) Name an international agency/organization for monitoring of radio spectrum usage worldwide.
    (c) Name radio spectrum regulatory agencies in the US and Europe.
    (d) What are the issues of consideration under the purview of regulation?

2

WIRELESS WIDE AREA NETWORK (CELLULAR NETWORK)

Since long, people have sought to communicate on a worldwide basis with each other. It is no more a privileged service; rather it is a bare need. Worldwide communication has become possible due to technological advancements in diverse fields of telephony and the Internet. The dream of communication, even in geographically remote areas, has come true through the introduction of wireless networks.

The most popular way to categorize communication networks is based on their scale—that is, the magnitude of area coverage. The Wide Area Network (WAN) covers a large geographical area. The Internet, spanning the earth, is the most prominent example of WAN in the wired network domain. In the wireless domain, the major example of WAN is the cellular network. This chapter deals with the cellular network and investigates the internals of a wireless WAN.

2.1 THE CELLULAR CONCEPT

Traditionally, the WAN is a collection of Local Area Networks (LANs) dispersed geographically. The relatively new cellular network is the collection of small networks, operated by the service provider(s) local to the networks. This concept was introduced in the early 1970s at Bell Laboratories. One of the most successful initial implementations of cellular concept was the Advanced Mobile Phone System (AMPS). It has been popular in the United States since 1983.

The system capacity of a cellular mobile network can be interpreted as the maximum number of users that can be supported at a particular point of time. Support to a large number of users leads to a large area

coverage. If a single transmitter is used to cover a large geographical area, a very powerful transmitter/antenna has to be installed. However, in reality, the system capacity offered by a single high power transmitter with full set of radio frequencies, allocated to the system, can't extend beyond a limit. Therefore, the alternative is to use a set of radio frequencies to serve a comparatively smaller geographical area and then reuse it for serving the other small areas. It avoids installation of the high power transmitter while keeping the desired system capacity. Care must be taken to ensure that the same set of frequencies, used to serve more than one geographical area, does not introduce interference among signals from the users in two different areas.

Partitioning a large coverage area with the target to provide small contiguous areas, supported by a low power transmitter with low antenna height in each of the areas, is the basis of a cellular communication network. The implementation of such a system follows the concept of cellular radios.

The next subsection defines the components of a cellular mobile network.

2.1.1 The Components

The basic components of a cellular mobile network developed around the cellular concept are as follows.

Transceiver: The device capable of simultaneously transmitting and receiving radio signals is called a transceiver.

Mobile Station (MS): Mobile station is an electronic device used either as a portable hand-held unit or mounted in a vehicle for communicating voice and data. It contains a transceiver, antenna and control circuitry.

Base Station (BS): A base station is a fixed station in the mobile radio system and is used for radio communication with mobile station (MS). It is normally located at the centre (Figure 2.1) of a network cell. Alternatively, it may also be placed on the edge of a coverage region. A base station contains transmitter, receiver, control unit, and antenna; and it has radio channels for allocation.

Mobile Switching Centre (MSC): An MSC coordinates routing of calls in a large service area. It connects both the base station and mobile stations to a telephone network called Public Switch Telephone Network (PSTN). That is, the MSC is an interface between the radio system and PSTN. The MSC is also referred to as the Mobile Telephone Switching Office (MTSO).

Forward Voice Channel (FVC) or downlink channel: The channels used for voice transmission from a base station to the mobile station are called FVC or downlink channel.

Reverse Voice Channel (RVC) or uplink channel: The RVCs or uplink channels are used for voice transmission from a mobile station to the base station.

Forward Control Channel (FCC)/Reverse Control Channel (RCC): These two channels (FCC and RCC) are responsible for initiation of mobile calls and are often called set-up channels. The FCC continuously broadcasts all traffic requests targeted to the MSs and serves as beacon. In general, on an average 5% of the total available channels within a system are assigned as the control or set-up channels.

2.1.2 Cellular Architecture

Figure 2.1 illustrates cellular network architecture with its components, the MS, BS and MSC. The whole geographical area covered by the network is partitioned into a number of hexagonal cells. Each such cell is serviced by a BS and a set of BSs are connected to an MSC. In addition to its function as a switching exchange, an MSC handles/controls (mobility management) of MSs in coordination with the BSs. The MSCs of the network are connected to the PSTN through a wired connection. The only wireless connection in a cellular network is in between an MS and the BSs.) — wireless.

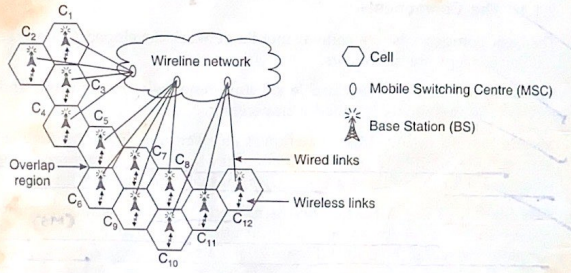


Figure 2.1 A cellular mobile network architecture.

The hexagonal cell shape considered in Figure 2.1 is purely imaginary. In reality, no such cell shape exists in the cellular network implementation. However, for proper planning of the system architecture, the hexagonal shape is universally adopted for several reasons.

Cell shape: In a cellular network, the geographical coverage region is considered to be partitioned into a number of small areas called cells. These cells are hexagonal in shape. The actual radio coverage of an antenna installed within a network cell is called the footprint. Although the real footprint is amorphous, the assumption of regular cell shape, in cellular network, is essential for a systematic system design. The probable candidate regular shapes can be the square, circular, hexagonal, etc. In a square cell based design, all the adjacent cell centres cannot be at the

same distance. The normal practice of installing BS antenna at the centre of a cell, therefore, cannot ensure the equidistance among the antennas. The antennas placed at equidistant simplify the task of determining when to switch a user (MS) to an adjacent antenna and as well as to choose the right one.

Considering circular radiation pattern and free-space propagation of the omni-directional BS antenna, the circular cell shape is mostly desirable. However, the circular cell can't fully cover a geographical region without the overlapping areas. On the contrary, a hexagonal cell approximates the circular radiation pattern. The adjacent BS antennas placed at the centres of such hexagonal cells are equidistant. Moreover, for a given distance between the centre of a polygon (circle/square/hexagon) and the farthest points on its perimeter, the hexagon has the largest area coverage among the three shapes.

The above discussions point to the fact that a hexagonal cell shape is most desirable in cellular network to cover a given network service area (coverage area). However, the exact hexagonal shape of cells cannot be implemented due to topographical and some other practical limitations of installing base station at a desired site. In reality, the designers practice variations in cell shapes. The reasons for such variations are:

- Due to practical limitations, e.g. land dispute, location unsuitability, some of the BSs can't be placed at the desired point (centre of a hexagon).
- The antenna at BS has some directional bias, in spite of the fact that it is desired to be omni-directional.
- The signal propagation characteristics may vary at places. In some cases, the transmission path between a transmitter and the receiver can be the simple direct line of sight. In a number of situations, the transmission path may be obstructed severely by high-rise buildings, foliage and terrain.

2.2 CALL SET-UP

When a Mobile Station (MS) is turned on within the coverage of a network, the following tasks are executed:

(i) The MS scans the group of FCCs (Forward Control Channels) in search of the strongest BS signal. The BS from which the strongest signal is received works as serving BS of the call initiator MS.

(ii) It checks continuously the selected FCC to ensure that the signal is not below the usable level.

In a cellular network, the call may be initiated by an MS or by a land phone. The call initiation process followed by an MS is different from that of a land phone. The steps to be executed during a call initiation are summarized in the next two subsections.

2.2.1 Call Initiation by an MS

Despite variations in the characteristics of MSs, the following sequential steps are commonly performed while a call is initiated by an MS:

(i) A call request is sent from the MS to the serving BS on RCC (Reverse Control Channel). The request message contains the MIN (Mobile Identification Number), ESN (Electronic Serial Number) and phone number of the called party (MS/land phone).

(ii) On receiving the request message, the serving BS transfers the request to the concerned MSC (Mobile Switching Centre).

(iii) The MSC, on receiving the message, validates the request/MS.

(iv) After successful validation, the MSC instructs the BS to assign a pair of unused FVC and RVC to the caller MS.

(v) MSC connects the MS with the called party. If the called party is also an MS,
   (a) MSC dispatches the request to all its BSs.
   (b) The MIN of the called MS is paged by all the BSs and instructs the called MS to use the unassigned FVC and RVC pair.
   (c) The called MS responds over RCC.
   (d) The serving BS sends the acknowledgement to MSC.
   (e) The serving BS sends an alert message to the called MS to ring.

(vi) Both the caller and called MSs begin voice transmission and reception.

2.2.2 Call Initiation by a Land Phone

The call initiation from a land phone leads to the following steps of activities while connecting an MS.

(i) The MSC receives the call request from a land phone through PSTN.

(ii) The MSC connects the land phone with the called party (MS)
   (a) MSC dispatches the call request to all its BSs.
   (b) On receiving the call request message the MIN of called MS is broadcast by all the BSs.
   (c) The called MS responds over RCC.
   (d) The serving BS sends an acknowledgement to the MSC.
   (e) The serving BS sends an alert message to the called MS to ring.

(iii) On receiving the acknowledgement, MSC instructs the BS to assign a pair of unused FVC and RVC to the called MS.

(iv) The MS responds to the caller party (land phone) over RVC for voice transmission.

2.3 FREQUENCY REUSE AND CO-CHANNEL CELL

In a cellular mobile network, the system may have to support a large number of simultaneous calls from and to the MSs. It requires allocation of a large number of forward and reverse channel pairs, thereby requiring a wide frequency bandwidth, which is a scarce resource. The objective of frequency reuse in a cellular system is to employ the same frequency in more than one network cell. This reuse reduces the demand of such scarce resource. However, use of the same frequency in nearby cells may interfere preventing the desired information flow. Therefore, the design challenge is to determine the distance between two cells that can use the same set of frequencies within the acceptable limit of interference.

The frequency reuse ensures proper distribution of total available frequencies allocated to a network. The total allocated frequencies are partitioned into a number of distinct sets for distribution in different network cells. A group of adjacent cells, called the cluster or compact pattern, is then determined. Each cell of a cluster is then allocated a distinct set of frequencies. It implies that in a cluster/compact pattern there is no scope of frequency reuse. A frequency f is allocated to two different cells X_i and X_j ensures that X_i and X_j belong to two different compact patterns.

Co-channel cell: The cells that can use the same set of frequencies as assigned to a cell X are called the co-channel cells of X. In Figure 2.2, the set of cells $S = \{F_1, F_2, \ldots, F_{19}\}$ form the cluster/compact pattern. Each cell of S uses a unique set of frequencies from the total allocated frequencies

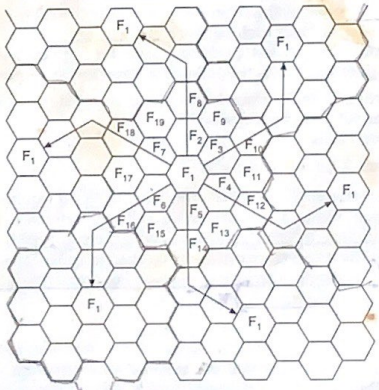


Figure 2.2 The co-channel cells.

of the network. The cells labeled with F_1 and are located outside the compact pattern are the nearest co-channel cells of the cell that is labeled with F_1 and is located at the centre of the compact pattern. The co-channel cells use the same set of frequencies for allocation to the MSs under their control. The minimum distance between the centres of any two co-channel cells that are nearest to each other is D. The distance D defines the effective measure of frequency reuse while avoiding any possibility of interference beyond the acceptable limit.

The repetition of compact pattern in the network coverage area places the co-channel cells in tiers. In general, a cell is surrounded by $6T$ co-channel cells in tier T. Figure 2.2 indicates 6 co-channel cells F_1 only at tier 1($T = 1$)—that is, the nearest co-channel cells. Similarly, tier 2 contains 12 co-channels. The co-channel cells of a network cell in each tier form a hexagon. The minimum distance between a cell (centre cell labeled with F_1) and its co-channel cell at tier T is TD.

The number of cells in a cluster $N = i^2 + ij + j^2$, where i and j are two positive integers. The size of cluster depends on the parameters i and j. In the figure $i = 3$, $j = 2$ and, therefore $N = 19$. The N is also considered as the measure of frequency reuse factor. In TDMA and FDMA, the frequency reuse factor is greater than 1 whereas in CDMA, this reuse factor is 1. In some literature $1/N$ is considered as reuse factor.

Example 2.1 Let us consider for a cellular system, 33 MHz bandwidth is allocated. Each of the full-duplex voice and control channels uses 50 kHz bandwidth. The number of channels can be allocated per cluster is, therefore, 33000/50 = 660. Now, if the cluster size chosen is $N = 12$ ($i = 2$, $j = 2$), for a cluster of size 12, then each network cell gets 660/12 = 55 channels.

To identify the nearest co-channel cells for a given cell X, and for particular values of i and j, the following steps are to be performed:

- Starting from X, move i cells along any chain of hexagons
- Turn 60° counterclockwise
- Move j cells

In Figure 2.2, $i = 3$ and $j = 2$.

Co-channel interference: A signal other than the signal required for desired communication within a network cell is called interference. Interference creates disturbance in communication and affects adversely the system performance. The interference caused by co-channel cell is called co-channel interference. Co-channel interference plays an important role in determining the quality of service (QoS)[1].

[1] The QoS is measured by several parameters such as coverage, call blocking (unable to initiate a call), call dropping (call is stopped during conversation), audio quality (cross-talk), etc. The audio quality is mainly affected by co-channel interference.

2.4 CELL DESIGN

The allocated frequency spectrum in a cellular network is very limited. The maximization of system capacity, keeping a reasonable quality of service, under the constraint of limited spectrum is the measure of performance of a system. However, the system capacity and quality of service are the two conflicting requirements. Thus while selecting the cell and cluster size, these two constraints are to be considered.

For example, let us assume that a cellular system has S duplex channels and a group of K channels are allocated to each cell of the network. If N cells are in a cluster, then $S = KN$. If a cluster is replicated M times, the total number of allocated duplex channels is $C = MKN = MS$, out of which S channels are unique. Here C denotes the capacity of the network. It signifies that the capacity is directly proportional to the number of times a cluster is replicated. Figure 2.3 shows the two parameters D and R (R is radius of a cell) for $N = 7$.

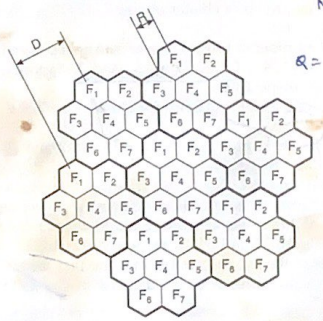


Figure 2.3 7-cell reuse pattern.

The ratio $Q = D/R$ is referred to as the frequency reuse ratio. From hexagonal cell geometry, the relation between frequency reuse ratio (Q) and the number of network cells per cluster (N) is

$$Q = \frac{D}{R} = \sqrt{3N} \tag{2.1}$$

The number of cells in a cluster (cluster size) is $N \propto D/R$. Now for a fixed cell size (R), if the cluster size (N) is increased, then D is to be increased. Alternatively, a large cluster size indicates that D/R is large. On the other hand, a network with small cluster size indicates its co-channel cells are located closer to each other and hence co-channel interference is significant.

The above discussions set the design criteria of a cellular network for the given coverage. These are (i) to maximize capacity ($C = M \times K \times N$), the smallest possible value of N is to be chosen and (ii) to reduce interference, the value of N—that is, the cluster size, is to be large. Thus to meet these two conflicting requirements (i) and (ii), a trade-off between the capacity and interference is desirable.

Example 2.2 Consider a cellular system having 1596 duplex channels to cover 2310 sq. km. The area of each cell is 6 sq. km. Assume that we need to compute the following:

(i) System capacity (C) for cluster size 7, 12 and 19.
(ii) M refers to number of times the cluster is replicated to cover the entire network area.

The total allocated channels in the system is $S = 1596$.

For the cluster size $N = 7$, 12 and 19, the number of channels per cell (K) = 1596/7, 1596/12 and 1596/19 = 228, 133 and 84 respectively.

The coverage area of a cluster of size 7/12/19 is $7 \times 6/12 \times 6/19 \times 6$ = 42/72/114 sq. km.

That is, the number of times a cluster is to be replicated is M = 2310/42, 2310/72 and 2310/114 = 55, 32 (approx) and 20 (approx) for the cluster size 7, 12 and 19 respectively.

Therefore, the system capacity for cluster size 7 is $C = M \times K \times N$ = $55 \times 228 \times 7 = 87780$.

For cluster size 12 and 19, $C = 32 \times 133 \times 12$ and $20 \times 84 \times 19$, that is, 51072 and 31920 respectively.

The system capacity C = 87780 signifies that at any point of time maximum 87780 simultaneous calls are allowed in the network with cluster size 7. Similarly, if the cluster sizes are 12 and 19, the number of simultaneous calls permitted is 51072 and 31920 respectively.

The results of the above example indicate that the system capacity of a cellular system decreases with the increase in cluster size.

2.5 INTERFERENCE IN CELLULAR SYSTEM

Interference creates disturbance in communication and affects the performance of a cellular network. It causes impairments ranging from the decrease in system capacity to dropping of calls. Signal received by the base station (BS) from a communicating mobile station (MS) may encounter interference due to:

- transmission from other mobiles in the same cell
- transmission from other mobiles in the neighbouring cells
- background noise
- other neighbouring base stations operating in the same frequency band
- noncellular system leaks energy into the cellular frequency band

Some of the reasons for interference are the improper channel assignment strategies, co-channel interference and adjacent channel interference out of imperfect cell design, etc. The co-channel interference is described in Section 2.3.

Adjacent channel interference: Signal impairment to a frequency due to the presence of another signal of very close frequency is called 'adjacent channel interference'. It mainly occurs due to the imperfect receiver filters that allow close/adjacent frequencies to leak into the passband[2].

The adjacent channel interference can have a serious effect if two MSs (adjacent channel users), one is located very close to the BS and the other is at a distance from the BS, simultaneously transmit equal power. The BS receives more power from the MS close to it. The SNR (signal-to-noise ratio) for this communication is much higher than that for the MS at a distance. Such a case is referred to as the 'near-far' problem. Further, if the MS close to the BS transmits a signal with order of magnitude much higher than the MS at a distance, the SNR for the second MS may not be detectable. This effectively creates jams in the communication channel of a network.

2.5.1 Signal-to-Interference Ratio

The quality of a signal communicated is also measured by a parameter called S/I (signal-to-interference) ratio. To compute S/I, the following interference sources are ignored as a matter of convention:

- co-channel interference from the second and other higher tiers of the network
- adjacent channel interference (negligible in comparison to the co-channel interference)
- background noise

The inter cell interference is mainly dominated by the co-channel interference.

The average channel quality is a function of distance dependent path loss[3]. Effects of other issues of communication are normally ignored. If c be the number of co-channel interfering cells and I_i be the interference

[2] In digital communication, the frequency band is split up into two main parts—the baseband and the passband. The passband contains all frequencies above a limiting frequency whereas the baseband refers to the frequencies below the limiting frequency. In radio communication, a baseband signal must be converted to the passband signal for its successful transmission.

[3] Path loss is the reduction in power density of the radio signal as it propagates. It is one of the dominating elements that causes impairment in the propagation channel as well as distorts the information-carrying signal as it propagates over medium. As an MS moves away from its serving BS, the received signal at BS becomes weaker because of the reduction of power density of the signal due to the growing distance between the MS and the BS.

power caused by transmissions from ith interfering co-channel cell to the BS, the S/I at the desired mobile receiver is:

$$\frac{S}{I} = \frac{S}{\sum_{i=1}^{c} I_i} \tag{2.2}$$

Now, if r be the distance between the mobile station and serving BS, the desired received signal power S is proportional to r^{-k}, where k is the path loss component (in general, $2 \le k \le 5$). Further, if D_i is the distance between ith co-channel cell and the mobile station, the I_i is proportional to $(D_i)^{-k}$. Thus the received S/I at the mobile station can be expressed as:

$$\frac{S}{I} = \frac{r^{-k}}{\sum_{i=1}^{c} (D_i)^{-k}} \tag{2.3}$$

If only first-tier co-channel cells are considered, then $C = 6$, and when an MS is in cell boundary ($r = R$), then $D_i \approx D$, for $i = 1, \ldots, 6$. Therefore, from Eq. (2.3), it can be deduced that:

$$\frac{S}{I} = \frac{\left(\frac{D}{R}\right)^k}{6} = \frac{Q^k}{6} = \frac{(\sqrt{3N})^k}{6}$$

or,
$$Q = \left(6 \times \frac{S}{I}\right)^{1/k} \tag{2.4}$$

It implies, from Eq. (2.4), for a given acceptable signal-to-interference (S/I) ratio, the frequency reuse ratio Q can be computed. From a known Q, the cluster size/compact pattern can also be settled (Eq. (2.1)) for a network.

Example 2.3 Let us assume that for a GSM TDMA system, the acceptable S/I is 15 dB or more. If path loss exponent $k = 3$, the frequency reuse ratio $Q = (6 \times 10^{1.5})^{1/3} \approx 3.13$. Then the compact pattern size $N = Q^2/3 = 3.26 \approx 3$. That is, a 3-cell frequency reuse system is to be used with 15 dB S/I.

2.5.2 Interference Reduction

It is reported in the earlier section that the co-channel interference is reduced if the distance between co-channel cells is increased. However, the large distance among the co-channel cells increases the cluster size, thereby reduces the system capacity.

On the other hand, the adjacent channel interference can be reduced through careful filtering and the proper channel assignments. Since each cell of a network is given only a fraction of the total available channels, a cell need not be assigned channels that are adjacent in frequency. This enables provision for the greatest possible frequency separation between the channels to ensure reduced adjacent channel interference within a cell.

Controlling transmission power of mobile stations (MSs) is another effective option to reduce the overall signal-to-noise ratio caused due to reverse channel transmission. In such mechanism, the power levels transmitted by each mobile station are under constant control of the serving base station (BS) to ensure that each unit transmits the minimum power necessary to maintain a good quality link on the reverse channel. The transmission power control also helps to extend battery life and combats the 'near-far' problem. The MS closer to BS transmits less power compared to that of the MS at a distance while both the MSs are adjacent channel users. At the serving BS's receiver, it maintains almost the same SNR for all the MSs' transmitters.

2.6 CHANNEL ASSIGNMENT

At the time of installation of a cellular network, each cell is assigned a set of channels to provide services to the individual calls of the cell. The task of assigning frequency channels satisfying the frequency separation constraints is known as the channel assignment problem. As the use of mobile communication systems grows, more efficient channel assignment/allocation techniques are needed. The channels on a network are the scarce resources and its proper management is very much desirable. An efficient assignment is not only important for new call initiation but it has also a great impact on call management while a mobile station moves from a cell to another—that is, handoff.

Channel assignment strategies are classified as the fixed and dynamic assignment. In fixed assignment, each cell of the network is to be allocated a predetermined set of voice channels. To provide call service, the cell can only utilize channels from that set. Any new call attempt within the cell is served while considering the channels remain unused at that moment. If all the channels of that cell are occupied, the attempted call is blocked.

One of the variations of fixed channel assignment is the channel borrowing. In this strategy, a cell is allowed to borrow channels from a neighbouring cell when its channels are already occupied and a handoff/new call is to be serviced. The MSC supervises such borrowing processes ensuring that this does not disrupt or interfere with the calls in progress. That is, the channels currently not being used by the donor cell or by its co-channel cells can be borrowed.

In dynamic assignment, the voice channels allocation to cells are not predetermined. Each time a call request is made by an MS, the serving BS seeks a channel from the MSC. The MSC in turn allocates a channel to the cell, requesting for the channel. While assigning channels, an MSC should consider the following:

- Likelihood of future blocking within the cell
- Reuse distance
- Cost functions

A number of works are so far been reported on channel assignment strategies. The main objective of such works is to ensure maximum utilization of radio resource, keeping the call blocking probability as low as possible. Such a region/area is defined as the hotspot. Therefore, an effective channel assignment strategy has to take care of the hot-spot areas.

2.7 HANDOFF

During its move an MS (mobile station) may occasionally switch to a new cell of the network. Switching from a cell to its neighbouring cell implies change of BS (base station) as well as switching to new voice and control channels. The process of transferring an MS from the control of current serving BS to a new one is called handoff. That is, the handoff process includes identification of a new BS for the MS and also allocation of the voice and control channels associated with the new BS. Handoff must be performed successfully and as infrequently as possible to reduce overhead due to switching.

The early prediction of handoff is the requirement for an efficient handoff-handling scheme. The simplest implementation is that whenever the power level from an MS received by a BS crosses a threshold ($P_{rthreshold}$), the BS initiates the handoff process. However, the whole process should be transparent to the users. To realize it, the system designer must specify an optimal signal level ($P_{rthreshold}$) to initiate a handoff. This threshold signal level is defined as $P_{rthreshold} = P_{rminimum} + \Delta$, where $P_{rminimum}$ is the minimum usable signal for acceptable voice quality at BS receiver and Δ is a small value.

If $P_{rthreshold}$ is too large, there may be unnecessary handoffs. The unnecessary handoff is considered to be costly to an MSC. On the contrary, if $P_{rthreshold}$ is too small, the system may not get sufficient time to complete the handoff process. While the BS/MSC handling the handoff, the MS may move farther from the serving BS and the call may be lost due to weak signal. As the handoff is costly, prior to any handoff, the system has to ensure that the drop in the measured signal level at serving BS is not due to any other reason except the movement of the MS away from the serving BS. In reality, a serving BS monitors the signal level received from an MS for a certain period of time. If the MS moves very fast, the slope of the short-term average of received signal becomes steep indicating that the handoff is to be processed instantly.

2.7.1 Handoff Strategies

The handoff in a cellular network is a regular feature. A cellular network without handoff is of no practical use. Therefore, efficient strategies need to be devised to address this issue. The following handoff tackling strategies are considered to be effective in a cellular mobile network:

- Network-controlled handoff (NCHO)
- Mobile-assisted handoff (MAHO)

NCHO: In this technique, the BSs monitor the signal quality received from an MS and report it to the MSC. On receiving this information from the BSs, the MSC selects the new serving BS and initiates handoff. The new BS takes control of the MS.

MAHO: In mobile-assisted handoff, implemented in GSM, the process of handoff is jointly taken care of by the mobile station (MS) and the network. The MS measures the signal levels received by it from the various BSs using a periodic beacon, generated by the BSs, to keep track of the MSs. The MS then feeds the measured signal levels back to the MSC via the serving BS. The MSC finally takes the decision of handoff.

Based on the procedures followed to take care of the handoff issues in a cellular network, the handoff is classified as Hard and Soft handoff.

Hard handoff: In Time Division Multiple Access (TDMA) and Frequency Division Multiple Access (FDMA), the frequency reuse factor (N) is greater than one. The neighbouring cells within a cluster use different frequency bands of the available frequency spectrum in such an implementation. An MS in a cell X', receives signals from different cells (BSs) surrounding X'. When the signal strength received by an MS from a neighbouring cell (X) exceeds a predefined threshold, the MS is instructed to switch to it (X). A new frequency band is then allocated to the cell X to support the ongoing communication of the MS. Such a termination of the existing connection (with X') to establish a new connection (to X) is referred to as the Hard Handoff. In hard handoff, an MS can have the radio link only with a single base station at any point of time and, therefore, it is also known as the 'break before make'.

Soft handoff: In soft handoff, the new connection between an MS and the new BS is established prior to the termination of existing link between the MS and the currently serving BS. Unlike hard handoff, during soft handoff, an MS can simultaneously communicate to more than one BS. The Code Division Multiple Access(CDMA) systems implement soft handoff techniques. As in CDMA systems for the cluster size $N = 1$ (Section 2.3), the spatial separation of frequency spectrum for allocating separate frequency band in cells is not necessary. Further, the interference from neighbouring cells is not too serious in such systems. Therefore, it can allow an MS to be connected to more than one BS simultaneously. If the MS enters a region in which the transmissions from two BSs are comparable, the MS then gets connected to the two BSs. The MS remains connected to the two BSs until one BS clearly predominates and the MS is assigned to that cell.

1st generation vs 2nd generation handoff

In 1st generation cellular systems, the voice transmission is analog. It uses FDMA technique to share the available spectrum. As cellular mobile systems have grown from analog to digital, the techniques of handling handoff get

the maturity. In 1st generation analog cellular systems, the notable features of handoff processing techniques are:

- Handoff decisions are taken by MSC with assistance of BS
  - BS measures signal strength (SS)
  - MSC supervises
- Each BS constantly monitors SS of RVCs (reverse voice channels) and determines the relative location of each MS

On the other hand, in 2nd generation digital cellular systems, which use TDMA technology, the following are the important features for implementing handoff:

- Handoff decisions are mobile assisted (MAHO)
- MS measures the power levels received from the BSs of neighbouring cells
- The measured values are continually sent to the serving BS

In MAHO, handover of calls between the BSs is much faster than that realized in 1st generation systems. An MS continuously measures the parameter values required for the handoff. The MSC no longer remains constantly busy for monitoring the signal strength at the MS. MAHO is particularly suitable for micro cellular (to be discussed in Section 2.7.2) environments where handoffs are more frequent.

Prioritizing handoffs

From the user's point of view, abrupt termination of a connection/call during running conversation is more annoying than a call is blocked occasionally on a new call attempt. Hence prioritizing handoff over the new call attempt is desirable. That is, at an instant of time if the system has to support a handoff and a new call attempt simultaneously, then a strategy is to be adopted so that service to handoff is considered first. There are two standard handoff strategies—the guard channel concept and the queuing of handoff requests.

In the guard channel, a fraction of the total available channels within a network cell is reserved for providing service to the handoff requests from ongoing calls. It means a cell gives priority to the handoff calls. On the other hand, in the queuing based handoff handling technique, there is no such reserved channel for the handoff calls. At the time of handoff if there is no available channel to serve, the call request is queued. Therefore, it decreases probability of forced termination of a call but delays the handoff.

The handoff strategies, guard channel concept and queuing of handoff, limit the total carried traffic in a network. In guard channel based technique a comparatively fewer number of channels can be allocated for the new call attempts by the MSs as a set of channels is kept aside for handoff calls. This reduces the total carried traffic within the system at any instant of time. On the other hand, in queuing the probability of a successful handoff improves at the cost of increased blocking probability of originating calls.

2.7.2 Constraints

The practical constraints in performing handoff are the accommodating users (MSs) with different speeds and the cell dragging.

Users with different speeds

In the cellular network, the mobile users can be inside a high-speed vehicle or they can be simply pedestrians. For pedestrians carrying MSs, the handoffs do not take place very frequently. So, the cell size in the network can be kept smaller (micro cell approach) without any severe problem in processing handoff. On the other hand, for fast moving users (sitting in a high speed vehicle) micro cell approach is not desirable. Micro cell based design for such cases may result in frequent events of handoff, thereby, increases computational overhead at the system components e.g. BS, MSC, etc. For high-speed mobile stations, the cell size is to be large (macro cell approach) to ensure efficient handling of handoff avoiding the excess loads in system components.

Theoretically, the cellular concept provides additional system capacity through addition of new cell sites. However, in reality, it is difficult for the service providers to identify new cell sites in urban areas even for some non-technical reasons. The practical solution is to use different antenna heights as well as the different power levels in the same location, to enable the design of 'micro' and 'macro' cells. This technique is called the 'umbrella cell' approach (Figure 2.4). The umbrella cell configuration consists of one large BS (macro cell) with high transmission power and antenna height. This set-up serves as an 'umbrella' for a number of small BSs (cells) with low transmission power and small diameters. While a high-speed user

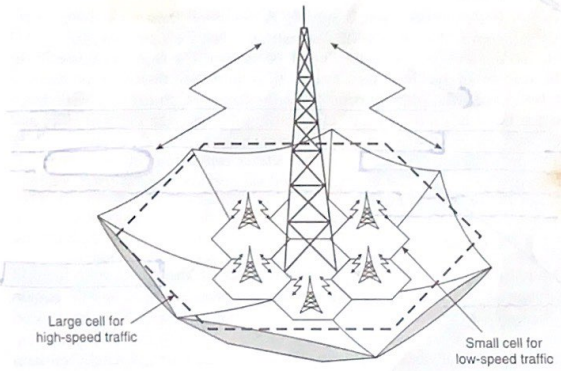


Figure 2.4 The umbrella cell.

(MS) within the large umbrella cell approaches the large BS, with rapidly decreasing speed, the BS decides to handover the MS into one of the co-located small BSs (micro cells) without any intervention of the MSC.

In GSM, the speed of a user can be determined almost accurately by the change of value of the parameter, timing advance (TA).[4] Its value is updated in the base station controller (BSC) located between BS and the MSC every 480 ms.

The umbrella cell based system (service area is covered by co-located micro and macro cells) has the following advantages over a microcell (service area is covered by microcells only) based system:

- Since it is not necessary to cover the whole service area with microcells, infrastructure cost is saved.
- The number of handoffs is much less than in microcell based system. In such a system whenever an MS, irrespective of its speed, crosses a cell boundary, a handoff occurs. On the contrary, in an umbrella cell environment, MSs with higher speed are under control of the BS of macrocell and the MSs with lower speed are under the control of the BS of microcell. So the MSs under macrocell are subject to less number of handoffs in comparison to the MSs under a microcell. Total number of handoffs is reduced in this way.

Frequent handoffs result in a substantial increase of the signalling load for a network. It degrades the quality of signal at the end user. The umbrella cell based system eliminates the possibility of frequent handoff for high speed MSs and very much effective in urban environments that feature city highways.

Cell dragging

In microcell environment, a serving base station may sense very strong signal from a slow-speed MS (pedestrian) even the user carrying the MS is well beyond the designated range of the cell. The signal received at the BS may be higher than the handoff threshold and, therefore, no handoff takes place. This phenomenon typically occurs in an urban environment when there is a line-of-sight (LOS) radio path between the user (MS) and the serving BS. It also causes interference when the MS penetrates deep into an adjacent cell area. Such a state is called the cell dragging. The parameter $P_{\text{rthreshold}}$ for taking handoff decision is to be chosen carefully to get rid of this problem.

2.7.3 Roaming

The roaming is a mechanism by means of which the intersystem handoff is taken care of. While the signal strength received by a mobile station (MS) becomes weak and the mobile switching centre (MSC) cannot find

[4] In GSM, timing advance (TA) corresponds to the time a signal from the MS takes to reach to the base station.

a cell within its system to handover the MS, it is handed off to a cell under the control of a different MSC. The intersystem handoff is called roaming (Figure 2.5).

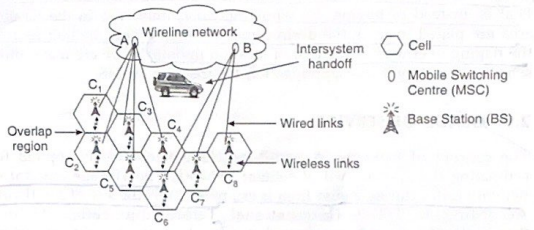


Figure 2.5 Intersystem handoff.

The important features of roaming are

- The MS moves out of its home network (to be described in Section 2.8).
- Local call becomes long distance call.
- Compatibility between two MSCs must be determined before implementing an intersystem handoff.

2.8 MOBILITY MANAGEMENT

The handoff management schemes are described in the earlier section. It is a part of mobility management in a cellular mobile network. The other important issue in mobility management is the location management. In this section, we concentrate on location management of a mobile station (MS) to keep track of its location for establishing seamless connection.

To start with an MS is assigned to a network called the home network of the MS. The association between the MS and the home network is made through a registration process. The home network is always aware of the current location of MS. While roaming from its home to a different network, an MS must register at the visiting network. Simultaneously, this update must reach its home network through the visiting network establishing an association between the home and the visiting networks.

A cellular network maintains two databases—Home Location Register (HLR) and Visiting Location Register (VLR). The HLR keeps the profile of mobile station and authenticates the subscriber before any updates. The VLR, on the other hand, authenticates the visiting MS in consultation with the HLR. The whole process which keeps track of a mobile subscriber's current location is called location management.

When a call/message destined to an MS is generated, the whole network is paged by the MSC to track the MS. In an alternative implementation, every time a mobile station crosses a location area (a group of cells), it sends its current location to the network (location update). That is, instead of paging the whole network, only cells in the location area are paged to track the desired mobile station. This method reduces the paging overhead at the cost of location update. There are many other schemes that target cost optimization for tracking an MS.

2.9 GRADE OF SERVICE

The concept of trunking in circuit-switching telephone is applied for estimating the system load of cellular network. The average amount of network traffic during a busy hour is estimated to fix the size of a network. According to ITU-T (International Telecommunication Union–Standardization Sector) recommendation, the busy hour is determined by considering the average of the busy hour traffic on the 30 busiest days of a year. Statistically, not all subscribers make calls at the same time and, therefore, it is reasonable to fix the size of network that can be able to handle expected level of load in the network. As a result of which whenever traffic load crosses the expected load, the probability of failure of an attempted call becomes non-zero and this may prohibit a user to access the trunked system. The ensured user access to a trunked system during the busiest hour is measured by the Grade of Service (GOS).

Cellular radio systems rely on trunking to accommodate a large number of users in a limited radio spectrum. There is a trade-off between the number of available pair of channels and the likelihood of a particular user finding that no channels are available during the peak time. C is the number of trunked channels offered (channel capacity) by a trunked radio system and A is the total offered traffic (offered load).

The trunking theory was developed by Danish mathematician Erlang. After his name a unit (Erlang) to measure the traffic intensity in a network is introduced. One Erlang represents the amount of traffic intensity carried by a channel when completely occupied. The traffic intensity A Erlang is expressed as $A = \gamma \times h$, where γ is the mean rate of connection requests attempted per unit time and h is the mean call holding time per successful connection. Thus A is the average number of calls arriving during the average holding period.

The traffic intensity A Erlang can also be expressed as per user traffic intensity (A_{pu}) and the number of users (n), i.e. $A = n \times A_{pu}$. If the average rate of connection request in a system is 25 calls/minute, and the average call holding time is 3 minutes, then the traffic intensity $A = 75$ (as $A = \gamma \times h$). This indicates if the channel capacity is exactly 75, it can meet the average demand. To meet the worst-case traffic load, the chosen capacity should be greater than 75. On the other hand, the channel capacity 150 indicates the channels in the system are remaining half-utilised.

In the trunked cellular system, when a user tries to communicate and all the channels are busy, then either the user is blocked, i.e. denied access to the system, or the user is queued until a channel is available. In the first case, the trunking is called blocked call cleared or lost call cleared (LCC); whereas the queuing is referred to as blocked call delayed or lost calls delayed (LCD). In general the cellular systems implement LCC. This model assumes the following:

- fixed arrival rate that follows poisson distribution
- memory-less arrivals of call requests, i.e. all users including blocked users may request a channel
- probability of a user occupying a channel is exponentially distributed so that long-duration calls are less likely to occur
- finite number of channels
- infinite number of users

An LCC model based system follows Erlang B formula to determine the blocking probability (P) of a new attempted call, where

$$P = \frac{\frac{A^C}{C!}}{\sum_{x=0}^{C} \frac{A^x}{x!}}$$

is the measure of GOS. For example, GOS = 0.001 implies that during a busy hour, on an average, one out of 1000 call requests may be blocked. Table 2.1 presents sample values of offered load to achieve the desired GOS for different number of channels.

Table 2.1 Capacity of an Erlang B System

Number of channels (C)	*Offered load*				
	GOS = 0.001	GOS = 0.002	GOS = 0.005	GOS = 0.01	GOS = 0.02
10	3.09	3.43	3.96	4.46	5.08
20	9.41	10.07	11.10	12.03	13.18
30	16.68	17.61	19.03	20.34	21.93
40	24.44	25.60	27.38	29.01	30.99
50	32.51	33.88	35.98	37.90	40.25
60	40.79	42.35	44.76	46.95	49.64
70	49.24	50.98	53.66	56.11	59.13
80	57.81	59.72	62.67	65.36	68.69
90	66.48	68.56	71.75	74.68	78.30
100	75.24	77.47	80.91	84.06	87.97

The table enables computation of—

- the GOS (P) for a given offered load (A) and the number of channels (C).

- the traffic load (A) that can be handled for a given number of channels (C) to achieve a desired GOS (P).

Example 2.4 Let us consider the network with $C = 20$ (number of channels/cell) and on an average each user makes 3 calls/hr. The average duration of a call is 2 minutes. The number of users supported in a cell with 1% blocking (GOS = 0.01) can be estimated from Table 2.1.

Now, as $C = 20$ and $P = 0.01$, then from Table 2.1, $A = 12.03$. Moreover, $A_{pu} = \lambda \times h = 3 \times (2/60) = 0.1$ Erlang. So, the number of users admissible to a cell = $12.03/0.1 \approx 120$.

Example 2.5 In a system with $C = 120$ channels, if the utilization of channel is 2/3, then the traffic intensity (load) in this network is $A = 2/3 \times 120 = 80$.

2.10 CAPACITY IMPROVING METHODS

The overall system capacity in a cellular network increases with the introduction of frequency reuse (Section 2.3). However, the frequency reuse cannot always avoid the cases of hotspot in a network. The channels assigned to the network cells may not be sufficient to accommodate all the users seeking connection. A set of cells may get over-committed resulting the hotspots (cells) within the network. This problem is taken care of by the technique called 'load balancing'.

To address the issue of hotspot, the capacity of a cellular mobile network can be increased simply by allocating/adding new channels. Alternatively, cells in the hotspot zone may borrow channels from the relatively cold cells (if any). However, a borrowed channel may be in use within the co-channel cell(s) of the lender cell. This causes channel interference and creates additional disturbances in information exchange. Therefore, during channel borrowing the necessary condition to be checked is that the lending cell has at least one free channel and that is not being used by any of its co-channel cell. Once borrowed, the channel in the lender as well as in all its co-channel cells are locked. Many variations of channel borrowing techniques are reported in the literature. However, it is still an active area of research. Detailed discussion on that is beyond the scope of this book.

Besides the manipulation of radio resources (such as channel), the other useful measures to increase the capacity of a system are 'cell splitting' and 'cell sectoring'. In 'cell splitting', the network cells are partitioned into smaller cells, whereas in 'cell sectoring', a cell is divided into several sectors resulting in increase in the system capacity.

2.10.1 Cell Splitting

The cell splitting is a mechanism of partitioning a network cell into a desired number of 'microcells', each having its own base station. When a cell becomes congested, it is partitioned into smaller subcells (microcells), as illustrated in Figure 2.6, to increase the system capacity. The heavy-traffic regions (e.g. hotspots) are normally split into such smaller areas to ensure the acceptable grade of service.

In a microcell, antenna height as well as the transmitter power is lowered. While a cell diameter ($2R$) ranges from 2 to 20 kilometres, the microcell diameter ranges from hundred metres to a kilometre. The cell splitting technique decreases R (Figure 2.6) while leaving Q (frequency reuse ratio) relatively unchanged. This mechanism allows a system to grow by replacing large cells with smaller cells without upsetting the channel allocation scheme required to maintain the minimum frequency reuse ratio (Q or D/R).

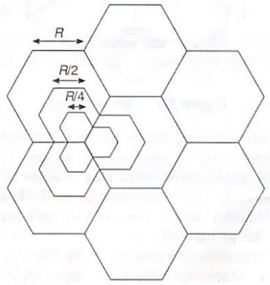


Figure 2.6 Cell splitting.

As an effect of cell splitting, the capacity of a network increases due to the additional number of channels per unit area. But the increased number of cells in a network adds more cell-boundaries. This may lead to more frequent handoffs. Therefore, the application of cell splitting is limited only to the cells with heavy traffic overloads. However, splitting only in a part of the system may result in serious channel assignment problems.

The designers favour cell splitting because of its capacity improving capability. The powers used by an MS and the BS after splitting is also comparatively low. The problem of unwanted handoff, however, is addressed with introduction of the concept of umbrella cells.

2.10.2 Cell Sectoring

Sectorization is the division of an omnidirectional (360°) view into non-overlapping slices called 'sectors' (Figure 2.7). Each sector has its own set of channels and considers directional antenna. The directional antennas at the BS are used to focus on the sectors. Usually this method divides a cell into three or six sectors. When a cell is sectored, the R (cell radius) is remained

unchanged. Further, D is reduced, and the amount of frequency reuse is increased. Hence the capacity is increased. Therefore, bandwidth efficiency of the system is enhanced as the frequency can be reused more often.

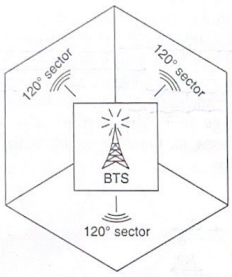


Figure 2.7 Cell sectoring.

The merit of cell sectoring lies on the fact that instead of interference received from all directions, it comes only from one direction causing the increase in signal to interference ratio. This enables to reduce the cluster size, and in effect allocates more channels per cell. However, within a cell an MS may need to handoff while crossing a sector boundary. The handoff is managed by the BS at the cell centre.

The cell sectoring adds complexity at the BSs due to introduction of additional antennas. Moreover, this mechanism decreases the trunking efficiency (queuing efficiency) while a large number of customers receive service from a set of servers. The proportional assignment of customers to different servers is a better solution.

2.11 USER VALIDATION IN CELLULAR COMMUNICATION

Just like any other network, user validation is an essential task in cellular network. In this section, the subscriber validation process, implemented in AMPS, is described.

2.11.1 Sources of Piracy

A common practice in unauthorized use of terminals is the cloning of stolen terminals. In AMPS, a terminal stores two ID numbers—the electronic serial number (ESN) and mobile identification number (MIN). The ESN is a 32-bit hardware-based serial number (unique 8-bit manufacturer's code + 24-bit ID for the mobile node for the given manufacturer's code). The MIN corresponds to a user telephone number (area code + phone ID)

assigned when a subscriber account is opened. Both the ESN and MIN are stored in the mobile terminal (MS).

The intruder can extract the ESN of a subscriber from the EEPROM (Electrically Erasable Programmable ROM) of the mobile device. The conventional measure for preventing unlawful access to such IDs is the encryption. The ESN is encrypted before writing them into the mobile terminal's EEPROM. Whenever the terminal tries to send ESN, the terminal decrypts the ESN first and then transmits over the air. That is, an ESN is easily accessible to the intruder who taps the network transmission. Therefore, it requires further elaborated security measure that prevents access to the network resources by an intruder. One such popular scheme is described in the next section.

2.11.2 Validation

The validation procedure in AMPS relies on the two IDs of a terminal (MS)—the ESN and MIN. The MS sends both the (ESN, MIN) to the network for:

- Registration
- Initiating a call
- Roaming

The subscriber validation process in the early AMPS system implemented verification of the transmitted tuple (ESN, MIN) with that already registered in the network. It also maintained the list of stolen terminal IDs to protect the system from piracy (Section 2.11.1). However, as both the IDs were transmitted straight over the air, an intruder had the scope to clone the ESN and MIN. Therefore, protection for unlawful access to the network was not guaranteed.

The advanced AMPS system uses a key-based authentication procedure to validate a subscriber. The serving network provides an A-key (64-bit primary secret key) value for each MS. During authentication a shared secret data (SSD, 128-bit shared secret key) is generated from the A-key and the ESN. The mobile terminal then runs an authentication algorithm and generates a terminal authentication result (MS-AUTHR). The MS-AUTHR along with the ESN and MIN is then transmitted from the MS to the network. Upon receipt of the information, the network registers the terminal. The serving network also computes SSD separately with the stored A-key and based on this SSD a network generated NT-AUTHR is produced. Finally, the NT-AUTHR and the MS-AUTHR are compared. If comparing a match indicates the MS is a valid user, the network allows the call initiated from the MS to proceed.

An unauthorised MS is not supposed to know A-key and hence the SSD is not known to it, the authentication data MS-AUTHR sent by the unauthorised MS will not match the NT-AUTHR. Therefore, the system should be able to recognise the unauthorised MS and denies its service.The

algorithmic steps of this authentication process (Figure 2.8) are noted below:

- An MS has a unique A-key supplied by the serving network.
- A SSD (shared secret key) is derived at the MS from A-key and ESN.
- MS runs an authentication algorithm and generates Authentication Result (MS-AUTHR) based on the SSD.
- The network produces a network generated AUTHR (NT-AUTHR) with the help of A-key stored in it.
- The NT-AUTHR and MS-AUTHR are compared at the network station. If there is a match, the network allows the call to proceed.

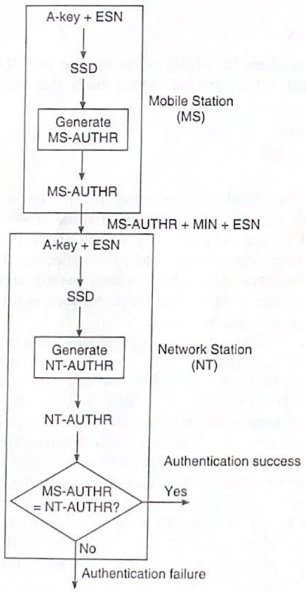


Figure 2.8 User validation process.

BIBLIOGRAPHY

Akeyama, A., T. Nagatsu and Y. Ebine, "Mobile Radio Propagation Characteristics and Radio Zone Design Method in Local Cities", *Review of the Electrical Communication Laboratories*, Vol. 30, No. 2, pp. 308–317, 1982.

Black, U., *Second-generation Mobile and Wireless Networks*, Prentice-Hall, New Jersey, 1999.

Blecher, F.H., "Advanced Mobile Phone Service", *IEEE Transactions on Vehicular Technology*, Vol. VT 29, pp. 238–244, 1980.

Boithias, L., *Radio Wave Propagation*, McGraw-Hill, Inc., New York, 1987.

Boucher, N., *Cellular Radio Handbook*, California, Quantum Publishing, 1991.

C.-L., I., L.J. Greenstein and R.D. Gitlin, "A Microcell/Macrocell Cellular Architecture for Low and High Mobility Wireless Users", *IEEE Vehicular Technology Transactions*, pp. 885–891, 1993.

Calhoun, G., *Digital Cellular Radio*, Boston, Artech House, Inc., 1988.

Cawthorne, N., *Cellular Radio—A European Round-up*, Wireless World, pp. 33–36, 1986.

Cox D.C., and D.O. Reudink, "Increasing Channel Occupancy in Large-Scale Mobile Radio Systems: Dynamic Channel Reassignment", *IEEE Transactions on Vehicular Technology*, Vol. VT-22, pp. 218–222, 1973.

Cox D.C., and D.O. Reudnik, "A comparison of some channel assignment strategies in large-scale mobile communications systems", *IEEE Transactions on Communications*, Vol. 20, pp. 190–195, 1972.

Cox D.C., and D.O. Reudnik, "Dynamic channel assignment in two-dimensional large-scale mobile radio systems", *Bell System Technical Journal*, Vol. 51, pp. 1611–1630, 1972.

Cox, D.C., "Wireless Network Access for Personal Communications", *IEEE Communications Magazine*, pp. 96–114, 1992.

Egli, J.J., "Radio Propagation above 40 MC over irregular terrain", *Proceedings of the IRE*, Vol. 45, pp. 1383–1391, 1957.

Everitt, D., "Traffic Engineering of the Radio Interface for Cellular Mobile Networks", *Proceedings of the IEEE*, Dayton, USA, Vol. 82, No. 9, pp. 1371–1382, 1994.

Halpren, S.W., "Reuse Petitioning in Cellular Systems", 33rd *IEEE Vehicular Technology Conference Record*, Toronto, Canada, 1983.

Kahwa, T.J., and N.D. Georganas, "A Hybrid Channel Assignment Scheme in Large-Scale Cellular-Structured Mobile Communication Systems", *IEEE Transactions on Communication*, Vol. COM-26, pp. 432–438, 1978.

Kozono, S., and M. Sakamoto, "Channel Interference Measurement in Mobile Radio Systems", *Proceedings*, 35th *IEEE Vehicular Technology Conference*, Agosto, pp. 60–66, 1985.

Lee, W.C.Y., "Elements of Cellular Mobile Radio Systems", *IEEE Transactions on Vehicular Technology*, Vol. 35, pp. 48–56, 1986.

Lee, William C.Y., *Mobile Cellular Telecommunications Analog and Digital Systems*, New Delhi, McGraw-Hill, Inc., International Editions, 1995.

Leonardo, E.J., and M.D. Yacoub, "Cell coverage area using statistical methods", *IEEE Global Telecommunications Conference (Globecom)*, Houston, USA, pp. 1227–1231, 1993.

MacDonald, V.H., "The Cellular Concept", *Bell System Technical Journal*, Vol. 58, No. 1, pp. 15–42, 1979.

Mark, J.W., and W. Zhuang, *Wireless Communications and Networking*, New Delhi, Prentice-Hall, 2003.

Mehrotra, A., *Cellular Radio: Analog and Digital Systems*, Boston, Artech House, Boston, 1994.

Oetting, J., "Cellular Mobile Radio—An Emerging Technology", *IEEE Communications Magazine*, Vol. 21, No. 8, pp. 10–15, 1983.

Pahlavan, K., and P. Krishnamurthy, *Principles of Wireless Network*, NJ, Prentice-Hall, 2002.

Pandya, Raj, *Mobile and Personal Communication Systems and Services*, New Delhi, Prentice-Hall of India, 2004.

Pollini, G., "Trends in Handover Design", *IEEE Communications Magazine*, 1996.

Rappaport, S.S., "Blocking, hand-off and traffic performance for cellular communication systems with mixed platforms", *IEEE Proceedings*, Vol. 140, No. 5, pp. 389–401, 1993.

Rappaport, Theodore S., *Wireless Communications Principles and Practice*, Prentice-Hall PTR, New Jersey, 1999.

Schiller, Jochen H., *Mobile Communications*, New Delhi, Pearson Education, 2007.

Sekiguchi, H., H. Ishikawa, M. Koyama and H. Sawada, "Techniques for Increasing Frequency Spectrum Utilization in Mobile Radio Communication Systems", 35th *IEEE Vehicular Technology Conference*, Agosto, pp. 26–31, 1985.

Siemens, "Telephone Traffic Theory and Table and Charts, Part 1", *Telephone and Switching Division, Siemens*, Munich, 1970.

Steele, R., J. Whitehead and W.C. Wong, "System aspects of cellular radio", *IEEE Communications Magazine*, Vol. 33, No. 1, pp. 80–86, 1995.

Tekinay, S., and B. Jabbari, "Handover and Channel Assignment in Mobile Cellular Networks", *IEEE Communications Magazine*, pp. 42–46, 1991.

Tripathi, N.D., "Handoff in cellular systems", *IEEE Personal Communications*, pp. 26–37, December 1998.

Whitehead, J.F., "Cellular Spectrum Efficiency via Reuse Planning", 35th *IEEE Vehicular Technology Conference*, Agosto, pp. 16–20, 1985.

Wong, D., and T.J. Lim, "Soft handoffs in CDMA mobile system", *IEEE Personal Communications*, Vol. 4, No. 6, pp. 6–17, 1997.

Young, W.R., "Advanced mobile phone service: introduction, background, and objectives", *The Bell System Technical Journal*, Vol. 58, No. 1, pp. 1–14, 1979.

REVIEW QUESTIONS

1. Define system capacity in the context of cellular mobile network.
2. What is cellular concept? How cellular concept enhances system capacity?
3. Describe cellular mobile network architecture. Which part of communication is wireless in such a network?
4. Explain the tasks of MS, BS, MSC, HLR and VLR.
5. What roles the channels FVC, RVC, FCC and RCC play in cellular mobile network?
6. Why hexagonal cell shape is normally used in cellular system design? In spite of the fact that hexagonal cell shape is most desirable, in practice, irregular cell shapes are considered. Why?
7. (a) Define frequency reuse. Why frequency reuse is essential?
   (b) What is co-channel cell?
   (c) How a compact pattern or cluster of cells is formed?
   (d) How the nearest co-channel cells are identified?
8. What are co-channel interference and adjacent channel interference? How they can be reduced?
9. What is meant by near-far problem? What mechanism is employed to combat such a problem?
10. Show that the system capacity increases with the number of times the cluster replicates.
11. (a) Consider 40 MHz bandwidth is assigned to a cellular mobile network. The system uses two simplex channels of 20 kHz to provide full-duplex voice and control channels. Calculate the number of channels to be assigned per cell for a cluster of size 12.
    (b) Consider a cellular system having 2023 duplex channels to cover 1925 sq. km and each cell area is 5 sq. km for 7-cell reuse system. Compute system capacity.
12. How the two parameters, system capacity and interference, play important roles in system design?
13. What is frequency reuse ratio? Derive the relationship between the frequency reuse ratio and signal-to-interference ratio.
14. When a BS receives signal from a communicating MS, state the causes of interference the signal may encounter.

15. Describe different channel assignment strategies.
16. Compare the following:
    (a) NCHO vs. MAHO
    (b) Hard handoff vs. soft handoff
    (c) 1st generation vs. 2nd generation handoff
17. Does MAHO require any action taken by the MSC? If yes, what role the MSC plays in such a scheme?
18. Why handoff prioritization is essential? How is it implemented?
19. What is umbrella cell? Why the concept of umbrella cell is used?
20. What is roaming?
21. What are the issues involved in mobility management?
22. (a) Define GOS and Erlang.
    (b) What are Lost Calls Cleared (LCC) and Lost Calls Delayed (LCD)?
    (c) Write assumptions (if any) to implement LCC.
23. Without manipulating radio resources, what are the other mechanism to increase system capacity?
24. Write the full forms of ESN and MIN. How are they used for validating a subscriber?

3

CELLULAR NETWORK STANDARDS (GSM AND IS-95)

The cellular Mobile Systems of the early era were very much analog systems. These were mostly installed in vehicles and not at all prepared to be used as hand-held devices. The analog systems of that era had poor voice quality and suffered from interference of signals. Further, in analog system security of communication was found difficult to implement. These factors limited the massive use of mobile systems. The low-cost, hand-held mobile station with longer battery life, better quality of service, enhanced security was only possible after the introduction of digital communication systems. This chapter attempts to make the readers familiarized with the commercially available digital cellular communication standards such as GSM, IS-95.

3.1 DIGITAL CELLULAR COMMUNICATION

Digital radio was first introduced in the defence sector to ensure quality reception with a high level of security in an interference-prone zone. The demand of hand-held terminal with reduced size and power requirement, longer battery life, etc. has led to explore digital cellular system commercially. An improvement in VLSI technology coupled with the advancement in the Digital Signal Processing (DSP) has made it a reality. The digital cellular system outperforms an analog system in terms of:

- capacity
- quality of service
- security
- ability to support improved services such as wireless Internet
- battery life

Digital scrambled up the signals into bursts; so it is more secure than analog and, therefore, helps prevention of stealing phone account information. However, roaming is comparatively difficult with a digital phone than an analog. As there is no uniquely accepted industry standard in digital technology, roaming, i.e. using an operator's network other than the home network can be difficult. On the other hand, an analog system has better coverage and the initial cost for analog is usually less than digital. Therefore, to get the best voice quality as well as security, the terminal offering dual mode (digital/analog) feature can be a better choice. This allows automatic switching between the analog and digital modes depending on the controlling antennae of the region the mobile station (MS) currently belongs to.

3.2 MULTIPLE ACCESS TECHNIQUES

The frequency spectrum allocated to an application (e.g. cellular communication) is shared by the users in a digital system. The sharing mechanism is realized with any one of the multiple access techniques mentioned in Section 1.2.3.

In an effort to make the most efficient use of radio frequency spectrum, a finite natural resource, various technologies have been developed. The target is to simultaneously support as many users as possible by a finite range of spectrum. In a cellular network, transmission from the base station (BS) in forward (downlink) channel can be received by all the MSs under its control. This, in a word, is broadcasting. On the other hand, in the reverse (uplink) channel multiple users want to transmit information simultaneously to the serving base station (BS). Without proper coordination among the transmitting users, collisions occur. Therefore, when two or more users transmit simultaneously to the same serving base station, it needs to be conflict free. This necessitates the scheme for multiple accesses.

The key target in multiple accesses is to separate the transmitted signals from different users at the base station. Different techniques of user separation at receiver (BS) have given birth to different multiple access techniques, namely Frequency Division Multiple Access (FDMA), Time Division Multiple Access (TDMA) and Code Division Multiple Access (CDMA).

In FDMA, a user separation is implemented by providing separate frequency band to individual user whereas in TDMA, it is done by providing separate time slot to each user. On the other hand, CDMA uses separate code to differentiate transmitted signal from an individual user. The 1st generation wireless systems use FDMA while the 2nd generation systems use TDMA and CDMA.

3.2.1 FDMA

An FDMA system separates out the total available bandwidth into several non-overlapping smaller bands/channels. Each channel has the ability to support one user. Each user is assigned a unique frequency band or channel.

Figure 3.1 illustrates that each user is allowed a unique frequency band. These channels are assigned on demand. When an MS tries to communicate with a BS, it registers itself using control channel to the closest BS. During registration, the BS assigns the MS an available pair of channels, one to transmit (reverse channel) and the other to receive (forward channel).

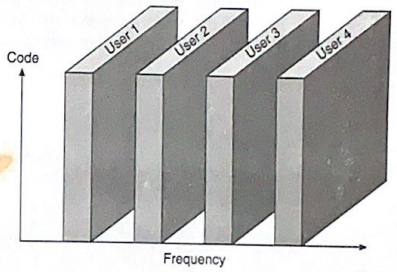


Figure 3.1 Frequency division multiple access.

To ensure interference-free transmission between two users or between the forward and reverse channel transmission of a user, it introduces the guard bands. The guard band separates out two such channels preventing interference. The guard bands are unused portions of the spectrum. For example, if 500 MHz bandwidth is provided for a system, entire bandwidth can be split into 12 channels each of 40 MHz. Each 40 MHz channel includes a 4 MHz guard channel. Effectively each channel is 36 MHz wide. In case of forward and reverse channels, either two antennas operating at different frequencies or one antenna with Frequency Division Duplexing (FDD) can provide the guard band.

As the multiple access is achieved by separating the user by the frequency allocated to them, frequency planning has to be done very carefully to avoid adjacent channel interference. This adjacent channel interference is an important factor in determining channel quality. However, frequency planning is complicated and difficult to achieve. Available frequency bands must be researched and analysed.

3.2.2 TDMA

In a system with a moderately large number of active users, the allocation of unique frequency band for each user cannot be a realistic option. As the demand increases to accommodate more users, spectrum efficiency[1] of

[1] Spectrum efficiency is measured as the amount of information that can be transmitted over a given bandwidth in a communication system. In the context of cellular mobile communication, spectrum efficiency means the maximum number of subscribers per cell that can be accommodated with an acceptable quality of service.

FDMA system becomes insufficient. A TDMA system splits users into an available pair of channels, but they also assign each user an available time slot within that channel. In each slot, only one user is allowed to either transmit or receive. Frequency division is still employed but these frequencies are now further subdivided into a defined number of time slots per frequency. Figure 3.2 shows that the entire spectrum is divided into four channels each of which is divided into three time slots. As the TDMA systems do not transmit all the time, the MSs gain an extended battery life as well as the talktime.

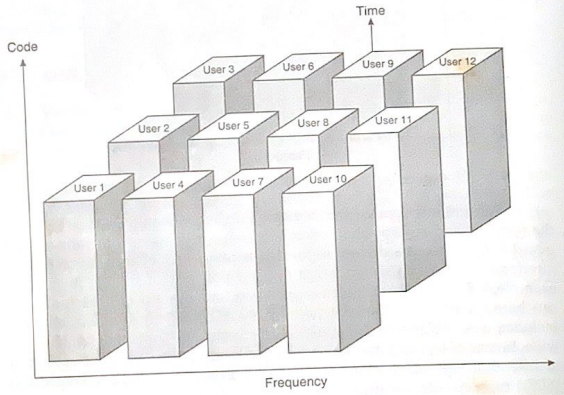


Figure 3.2 Time division multiple access.

In TDMA, like FDMA, the BS assigns the MS an available pair of channels. In addition to this, it assigns a time slot within the channel that must be available to the MS. The user can only be allowed to send or receive data for this time slot. The data bursts are fast reassembled at the receiving end, and therefore, appeared as continuous. Figures 3.1 and 3.2, point to the fact that while an FDMA system supports 4 users, the TDMA can support 12 users utilising the same bandwidth. As the TDMA systems first split an allotted portion of the frequency spectrum into smaller slots/channel, they require the same level of frequency planning as FDMA system.

3.2.3 CDMA

In code division multiple access (CDMA), signals from users (MSs) are differentiated at the receiver (BS) in code space whereas in FDMA it is

differentiated in frequency space and in TDMA, in time space. The main challenge of CDMA is to find suitable code for a user's signal so that it is unique and can be differentiated from the signals of other users. The hurdle is mitigated in CDMA by finding mutually orthogonal codes having high autocorrelation[2] value for each user. Two codes are orthogonal to each other when their cross correlation[3] value is 0. However, the CDMA does not allocate unique channel for a user. Each user can utilize the entire block of allocated spectrum space to carry the messages, that is, all the users can use the same carrier frequency and transmit simultaneously.

There are two popular coding techniques used in CDMA—walsh code[4] and PN (pseudo noise) code[5]. In the forward (BS to MSs) direction, transmission for all the users originates from the same transmitter (BS) in a perfectly coordinated manner. The orthogonal walsh code is used for forward channel transmission. In forward direction, transmission is originated from a BS, a fixed station. It enables the generation of orthogonal codes for the user signals. On the other hand, the reverse transmission (MS to BS) cannot be accurately coordinated due to the mobility of MSs. Generation of orthogonal codes for each user in such case is not at all possible for arbitrarily random locations of the MSs. Hence the PN code is used for the reverse channel transmission.

Spread spectrum

Spread spectrum is a technique that realizes the unique coding for signals from each user. This also improves the spectrum efficiency over that of FDMA and TDMA. The principle of spread spectrum communication is to spread the bandwidth of baseband information-carrying signals from different users to a much larger bandwidth. Ideally, the spreading signals used for different users are orthogonal to each other.

In spread spectrum, the narrow band information-carrying signal is multiplied by a very large bandwidth signal called the spreading signal. The bits of spreading signal are referred to as the chips. If T_c is the period of one chip, $1/T_c$ is the chip rate. The user data rate is $1/T_b$, where T_b is the period of one data bit (baseband information-carrying signal). The spreading factor is defined as the ratio of chip rate and user data rate (T_c/T_b).

One of the popular implementation of spread spectrum is the direct sequence spread spectrum (DSSS). In DSSS-based CDMA, the spreading signals with different chip rates are used to the different data signals. The chip rates used for the spreading signal are significantly greater than the

[2] Correlation is to determine how much similarity a code has with another code. It is computed by taking inner product of the codes. Autocorrelation is the correlation of a user code with itself.
[3] Cross correlation is correlation between two separately generated codes.
[4] Walsh codes are mathematically orthogonal codes.
[5] A PN code is a binary sequence that appears randomly but can be reproduced in a deterministic manner.

user data rates. Normally, the chip rate is in the order of "megachips per second" (Mcps), that is, millions of chips per second. For DSSS-based CDMA modulation (Refer to Figure 3.3(a)), the first task is to spread the information-carrying (user data) signal (digital modulation) and the second task is to modulate spread signal with carrier frequency. For example, spreading a user signal of 1 MHz bandwidth with 11 chip code generates a signal of 11-MHz bandwidth. Then the signal is converted to GHz bandwidth to make it ready for transmission. Demodulation (Refer to Figure 3.3(b)), the reverse process of modulation, is employed to retrieve user data at the receiver (BS) side.

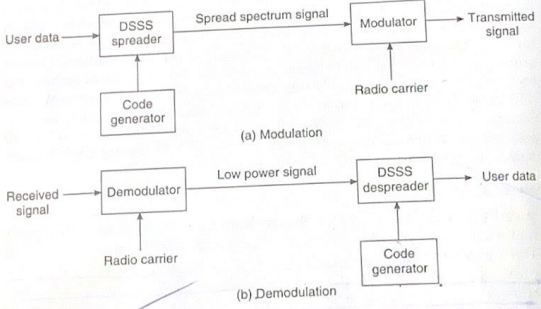


Figure 3.3 DSSS CDMA system.

If sufficient bandwidth is available, spreading is performed in two consecutive phases. In one phase data is spread by an orthogonal code to provide mutual orthogonality among all users in the same cell. In the next phase the resulting signal is further spread by a PN sequence to provide mutual orthogonality among the users in different cells. Such two phase spreading is called multiple spreading. IS-95 adopts this multiple spreading technique.

In CDMA for both the forward and reverse channel transmission, unique codes (walsh code and PN code respectively) are used for a user. A receiver would try to reconstruct only the specific desired user's data. And all other users' data would be sensed as noise. While detecting the user data signal, a receiver (BS) needs to know the code used by the transmitter (MS). An analogy can be drawn with a business party being held in a big hall. As it is shown in Figure 3.4, a number of persons are sharing views among themselves. Each pair of them is using a unique language for conversation. As a pair is using separate language (code), others conversations cannot be affected severely provided no one speaks much loudly compared to others. If one speaks so, it will add to noise.

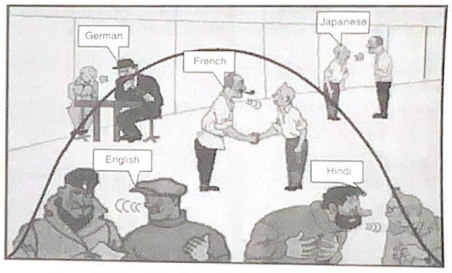


Figure 3.4 An analogy with CDMA.

It is mentioned earlier that the CDMA demands the codes should be orthogonal to each other and a code for a certain user should have a good autocorrelation. For example, let us consider the two simplest codes (1, 0) and (1, 1). Assuming 0 as –1 and 1 as +1, the codes become (+1, –1) and (+1, +1). The inner product of these two is $(+1) \cdot (+1) + (-1) \cdot (+1) = +1 - 1 = 0$ and hence the codes are orthogonal. A code with high autocorrelation means the absolute value of inner product of the code with itself is high. For example, for the code (+1, –1), the autocorrelation value is $(+1, -1) \cdot (+1, -1) = (+1) \cdot (+1) + (-1) \cdot (-1) = 1 + 1 = 2$. An example of a code, used for ISDN and IEEE 802.11 is 11 chip code (+1, –1, +1, +1, –1, +1, +1, +1, –1, –1, –1). The autocorrelation value of this code is 11.

How it works

Consider two MSs, MS_a and MS_b that want to transmit $data_a$ (1, 0) and $data_b$ (1, 1) respectively. CDMA assigns two unique and orthogonal codes—$code_a$ = (0, 1, 0, 1), $code_b$ = (0, 1, 1, 0). Both the MS_a and MS_b spread respective signal using their unique key as chip sequence (spreading signal). The spreading algorithm implies that if the data bit is 1, then send the corresponding code itself and for data bit 0, send the complement of the corresponding code. The resulting signals (encoded $data_a$ and $data_b$) are given below. For convenience binary 0 is assumed as –1, and binary 1 as +1.

Encode MS_a

$code_a = (-1, +1, -1, +1)$
$data_a = (+1, -1)$
encoded $data_a = code_a \cdot data_a$
$= (-1, +1, -1, +1) \cdot (+1, -1)$
$= (-1, +1, -1, +1), (+1, -1, +1, -1)$
$signal_a = (-1, +1, -1, +1, +1, -1, +1, -1)$

Encode MS_b

$code_b = (-1, +1, +1, -1)$
$data_b = (+1, +1)$
encoded $data_b = code_b \cdot data_b$
$= (-1, +1, +1, -1) \cdot (+1, +1)$
$= (-1, +1, +1, -1), (-1, +1, +1, -1)$
$signal_b = (-1, +1, +1, -1, -1, +1, +1, -1)$

The received signal at the BS is the sum of these two signals; $signal_a$ and $signal_b$, if there is no other MSs under the serving BS. That is, $signal_{received\text{-}at\text{-}BS} = signal_a + signal_b = (-2, +2, 0, 0, 0, 0, +2, -2)$. As the $code_a$ and $code_b$ are orthogonal to each other, the $data_a$ and $data_b$ can easily be extracted from the $signal_{received\text{-}at\text{-}BS}$. The decoding performed at BS to extract the data sent by the MS_a and MS_b are described below.

The extraction of data sent by MS_i means:

(i) $decode_i = signal_{received\text{-}at\text{-}BS} \cdot code_i$

(ii) replace values greater than 0 of $decode_a$ as 1 and values less than 0 as 0 to get the $data_i$. Therefore

Decode MS_a	Decode MS_b
$code_a = (-1, +1, -1, +1)$	$code_b = (-1, +1, +1, -1)$
$signal_{received\text{-}at\text{-}BS} = (-2, +2, 0, 0, 0, 0, +2, -2)$	$signal_{received\text{-}at\text{-}BS} = (-2, +2, 0, 0, 0, 0, +2, -2)$
(i) $decode_a = signal_{received\text{-}at\text{-}BS} \cdot code_a$	(i) $decode_b = signal_{received\text{-}at\text{-}BS} \cdot code_b$
$= ((-2, +2, 0, 0), (0, 0, +2, -2)) \cdot (-1, +1, -1, +1)$	$= ((-2, +2, 0, 0), (0, 0, +2, -2)) \cdot (-1, +1, +1, -1)$
$= (2 + 2 + 0 + 0), (0 + 0 - 2 - 2)$	$= (2 + 2 + 0 + 0), (0 + 0 + 2 + 2)$
$= (+4, -4)$	$= (+4, +4)$
$decode_a = (+4, -4)$	$decode_b = (+4, +4)$
(ii) $data_a = (1, 0) = (+1, -1)$	(ii) $data_b = (1, 1) = (+1, +1)$

A CDMA system provides more privacy than FDMA or TDMA systems. Due to the wide bandwidth of a spread-spectrum signal in CDMA, it is very difficult to cause jamming or interference. It appears as nothing more than a slight rise in the noise floor[6] or interference level. In other technologies, the power of signal is concentrated in a narrower band and, therefore, easier to detect or decode.

Increase in number of users in a CDMA system linearly raises the noise floor. Thus there is no absolute limit on the number of users in CDMA. Rather, the system performance gradually degrades for all users as the number of users is increased. Hence the CDMA has a soft capacity limit[7]. Moreover, there is a problem of self-jamming[8]. Self-jamming arises while spreading sequences of different users are not exactly orthogonal resulting in one user disrupting the transmission to the other, and vice versa. Further the users of CDMA share the same channel. It may lead to the near-far problem (Section 2.5). To combat this problem, power control

[6] Noise floor is the measure of sum of signal generated from all the noise sources.

[7] Soft capacity limit is a limit of accommodating number of users beyond which system performance degrades to an unacceptable extent.

[8] Transmission of radio signals that disrupts communications by decreasing the signal-to-noise ratio.

technique is employed in most CDMA implementations so that each MS under a BS coverage can provide the same signal level to the BS receiver.

The well-designed filters employed in FDMA ensure zero or a minimal spectral overlap whereas employment of slot synchronization in TDMA reduces timing jitter. This makes both FDMA and TDMA conflict-free multiple access schemes. On the other hand, CDMA is a spread spectrum technique. Although orthogonality between transmitted signals from different users is the key element of CDMA, in reality the generated wideband spreading functions are not truly orthogonal and cause some interference. Hence CDMA is interference limited multiple access strategy.

The following discussion summarizes the merits and demerits of the three multiple access techniques—FDMA, TDMA and CDMA.

FDMA

Merits

- simple to implement
- fairly efficient with a small population and when traffic is almost constant
- fewer bits are needed as overhead (synchronization, etc.) as compared to TDMA

Demerits

- frequency planning is difficult
- if a channel is not used by the designated user, it cannot be used by others to increase or share capacity—a wastage of resources

TDMA

Merits

- data transmission for users is not continuous leading to low battery consumption
- can accommodate much more users in the same spectrum space as compared to FDMA; capacity increases in high traffic areas
- can allocate different number of time slots per frame to different users. Thus bandwidth can be supplied on demand to different users based on priority.

Demerits

- as high synchronization overhead is required, receivers are to be synchronized for each data burst and, therefore, requires additional overheads in comparison to FDMA
- frequency guard bands lead to spectrum inefficiency
- security is much lower than that in CDMA
- frequency planning is critical
- the number of users accommodated is less than CDMA in the same spectrum space

CDMA

Merits

- best spectrum efficiency: capacity increases 8 to 10 times than that of an analog system and 4 to 5 times than the other digital systems, making it most useful in high traffic areas with a large number of users and limited spectrum.
- simplified frequency planning as all users utilize the same radio frequency spectrum
- random walsh codes enhance user privacy/security

Demerits

- base station equipment is expensive
- difficult to implement in comparison to FDMA and TDMA
- continuous power adjustment of MS is essential to avoid near-far problem
- self-jamming arises from the fact that the spreading sequences of different users are not exactly orthogonal

Legacy commercial telecommunication networks such as analog networks based on Advanced Mobile Phone System (AMPS) are built around the FDMA. Currently, there are several digital cellular standards available worldwide. The example of TDMA-based standards are Global System for Mobile (GSM), Digital Cellular Systems (DCS), etc. On the other hand, IS-95 is the CDMA-based standard. The details of GSM (TDMA) and IS-95 (CDMA) standards are provided in the following sections.

3.3 GSM

Global system for Mobile (GSM) is a 2nd generation cellular system. It was developed in the 1990s to solve the fragmentation problems of first cellular systems and introduced in Europe. By 1993, GSM and its technically equivalent offshoot, DCS (Digital Cellular System) 1800, were adopted outside Europe, too.

GSM is the world's first cellular system to specify digital modulation and network level architectures and services. It has two objectives—pan-European roaming and interaction with the ISDN. Roaming yields compatibility around Europe and interaction with the ISDN offers capability to extend the single-subscriber-line system to a multi-service system. System capacity was not a primary issue at the initial development phase, however, to keep pace with the rapid growth in cellular services, many revisions have been made to GSM.

3.3.1 System Architecture

The GSM consists of three major interconnected components—Base Station Subsystem (BSS), Network and Switching Subsystem (NSS) and Operation

Support Subsystem (OSS). A mobile station (MS) interacts with BSS through air interface. The important interfaces are Um, A_{bis} and A. Figure 3.5 illustrates the block diagram of GSM system.

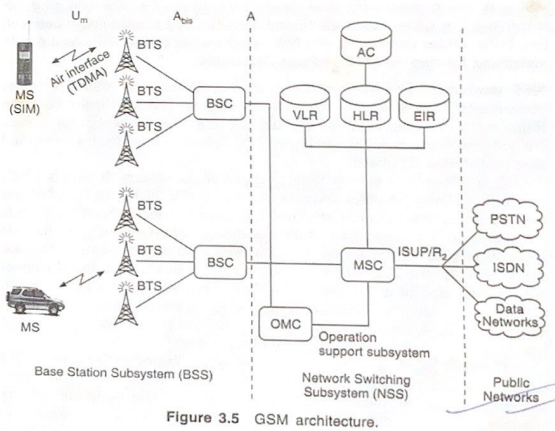


Figure 3.5 GSM architecture.

Mobile station: A mobile station comprises of device-dependent hardware/software and the subscriber-dependent SIM (subscriber identity module). The device is identified by IMEI (international mobile equipment identity) whereas a subscriber is identified by SIM. The device dependent hardware are transceiver, antenna, etc. and the softwares are for theft protection, etc. The SIM is a small detachable card and securely stores user specific information such as IMSI (international mobile subscriber identity), PIN (personal identity number), PUK (PIN unlocking key) and an authentication key to run user validation algorithm. The IMSI is similar to MIN whereas the IMEI is equivalent to ESN in AMPS. At any instant of time, an MS is either in idle state or in dedicated state. In the idle state, the MS only listens to the network but neither transmits nor receives any data. In this state it does not use any dedicated resources. On the other hand, in dedicated mode the MS transmits/receives data using dedicated resources. From an idle state, if an MS desires to access the network, it enters into the dedicated state and starts using resources.

BSS modules: A BSS consists of Base Station Controllers (BSCs), each of which controls a number of Base Transceiver Stations (BTSs). The BSCs

are connected to an MSC. An MS communicates BSS via the BTS through TDMA air interface (Figure 3.5).

A BSS provides radio transmission paths between an MS and the MSCs. It also manages the radio interface between the MSs and all other subsystems of GSM. Handoffs (Handovers in GSM terminology) between two BTSs, under the control of a BSC, are handled by the BSC. As a result, switching burden on MSC reduces drastically.

NSS modules: The MSC is the central component of NSS. The other components of NSS are Home Location Register (HLR), Visitor Location Register (VLR), Equipment Identity Register (EIR) and Authentication Centre (AC). There is logically one HLR per GSM network. It is also implemented as a distributed database.

The NSS manages switching function of the system. It allows MSCs to communicate with other networks such as PSTN, ISDN, etc. In addition, an NSS is responsible for all the functionalities needed to handle a mobile subscriber, such as registration, authentication, location updating, handovers and call routing to a roaming user. An MSC can connect data networks, e.g. X.25 with the help of IWF (inter-working functions), an additional module attached to it. However, the IWF is not shown in Figure 3.5. On the other hand, an MSC handles all signalling needed for connection set-up, release and handover to other MSCs using SS7[9] (standard signalling system No. 7) used in ISDN and current public networks.

The HLR is a huge database capable of managing to the tune of few million subscribers' data. It stores IMSI and corresponding list of subscribed services such as call forwarding, call waiting, roaming limitation, etc. In addition to this, the HLR maintains location information for each subscriber registered in the GSM network. If the system has to establish a call to an MS, the MSC seeks routing information from the HLR.

The VLR temporarily stores the IMSI and information for each subscriber, visiting the coverage area of an MSC. The VLR temporarily assigns TMSI (temporary mobile subscriber identity) to each roaming subscriber for concealing IMSI, a user identity. The VLR may change TMSI dynamically. The TMSI remains valid within the coverage of the VLR. A VLR is capable of managing up to one million subscribers. Although each functional entity can be implemented as an independent unit, most manufacturers of the switching equipment implement one VLR together with one MSC, so that the geographical area controlled by the MSC corresponds to that controlled by the VLR, simplifying the signalling required. When an MS becomes roaming that is, comes to a new MSC, it registers itself to the VLR of the MSC. Once it is registered is in the VLR, the MSC sends the necessary information to the visiting subscriber's HLR so that calls to the roaming mobile can be appropriately routed over the PSTN by the roaming user's HLR. The database in VLR is

[9] A set of telephony signalling protocols which are used to set up, maintain and release PSTN calls.

temporary, that is, the subscriber data is stored as long as the subscriber is within the service area. The HLR and VLR, together with the MSC, tackle the call routing and roaming facilities of GSM. Both the database (HLR and VLR) contains a large volume of data and, therefore, maintains suitable database organizations to retrieve the desired subscriber's information in real time.

Both the AC and EIR database are used for providing security. When EIR checks the validity of the equipment by checking IMEI, the AC provides information for verifying the subscribers SIM cards. The EIR database contains a list of all invalid MSs of the network. That is, the list contains the IMEI of MSs that are either to be banned or to be monitored. EIR marks an IMEI as invalid if the MS is stolen/blacklisted or its type is not approved. When such an MS seeks registration to a GSM network, the MS is requested to provide the IMEI for equipment verification. If the MS is found in the list, access to the network is denied to that MS. In some implementations, EIR is integrated with the HLR.

The AC is a database that keeps a copy of the secret-key stored in each subscriber's SIM card. When an MS keeps its power on, it attempts to connect to the GSM network. The SIM of the MS is authenticated at this stage. The AC combines the secret-key with IMSI to produce a challenge/response type of identification and supplies it to the MSC. If the SIM can generate the same identification with the secret-key and IMSI stored in it, the SIM can gain access to the network thereby authentication of SIM is done. This method of SIM authentication will be elaborated further during discussion on user validation in Section 3.3.8. In addition to SIM authentication, the AC participates in securing radio communication between MS and the network. Both the AC and the SIM produce a cipher key separately. With the available cipher key at both the MS and the network side, radio transmission from the MS is encrypted whereas at the network side it is decrypted and vice versa.

OSS modules: It supports one or more Operation Maintenance Centres (OMCs) and is solely accessed by the staff of GSM operating company. An OSS maintains all the telecommunication hardware and network operations within a particular coverage area. It is the manager of all sorts of charging and billing procedures and allows system engineers to monitor, diagnose and troubleshoot all the aspects of the GSM.

3.3.2 OSI Layers in GSM

The GSM is an OSI (Open System Interconnection) network as shown in Figure 3.6. It has three layers, namely physical layer (Layer 1), data link layer (Layer 2) and network layer (Layer 3). As the air interface between MS and BTS is provided by U_m, the protocol layers connected by U_m are given special attention. The other interfaces (A_{bis}, A, ISUP/R2) lie between the modules in a fixed network.

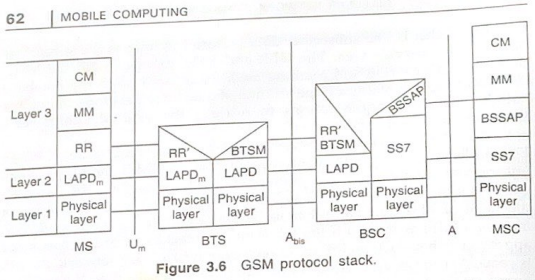


Figure 3.6 GSM protocol stack.

Physical layer: This layer is responsible for creation of bursts, multiplexing of bursts into a TDMA frame and finally for actual transmission of the data. The layer is responsible for channel coding and error detection/correction. Further, a physical layer includes some synchronization features that do not have any significance to the higher layers, since those features are purely hardware related. A physical layer is also responsible for detection of idle channels and the channel quality of downlink. However, it cannot identify the data types or data formats and cannot differentiate the control and user data. The data packets received from data link layer are transmitted without additional verification.

The physical layer at U_m interface performs encryption/decryption. The encryption in GSM is performed only between MS and BSS over the air interface. For the air interface of GSM systems, the GMSK (Gaussian Minimum Shift Keying) modulation is used. Over the terrestrial interfaces, i.e. in A_{bis} and A interfaces, data transmission at the physical layer uses pulse code modulation (PCM). The implementation of physical layer depends greatly on the type of interface.

Data link layer: This layer is responsible for the packaging of data. The data are combined into packets or frames and then handed to the physical layer for synchronous and asynchronous transmission. The main tasks of layer 2 are to detect and correct the errors. Data frames are formed by introducing start/stop marks and the check sums. When a receiver detects an error, it tries to correct the error or requests retransmission.

Layer 2 protocols might change from interface to interface. For example, layer 2 protocols for data exchange between an MS and the BTS is $LAPD_m$[10] (Link Access Procedure on the D channel). $LAPD_m$ is a modified version of the LAPD over U_m interface. Data exchange from the BTS to the BSC follows LAPD whereas data exchange between the BSC and the MSC follows MTP2/SS7 (message transfer part, layer 2 of SS7) as the layer 2 protocols.

[10] ITU (international Telecommunication Union) specifies the data link layer protocol LAPD in ISDN protocol stack on the D (data) channel referring to the ISDN channel that carries control and signalling information.

Network layer: It prescribes the path a message has to take and the recipient of that message. All the information necessary to route a data packet is provided at this layer (layer 3). A network layer is divided into three sublayers—Radio Resource (RR) management, Mobility Management (MM) and Call Management (CM) units. A part of RR called RR′ is implemented in BTS, and the remainder is installed in BSC. The functions of RR′ are controlled by the BSC via the BTS management (BTSM) unit. Most of the RR functions are performed at the BSS.

The time duration when a mobile is in dedicated mode and busy with the configuration of radio channels is called RR-session. The responsibility of the RR sublayer is to manage this session. The main tasks of this layer, therefore, are to set up, maintain, and release of radio channels. In addition to this, it manages power control, discontinuous transmission/reception and handovers that are elaborated below:

- *Power control:* It is to minimize co-channel interference and to conserve power level that maintains an acceptable signal quality at MS. Mobile Station decides the acceptable power level measuring the bit error ratio (BER)[11].
- *Discontinuous transmission:* This is a power-saving mechanism. It is possible to implement in GSM exploiting the fact that a person speaks less than 40% of time during a normal conversation. By turning off the transmitter for rest of the 60% time can save power. However, in order to distinguish voice and background noise, very accurate voice activity detector should be used. While transmitter is off, the receiving end will hear a total silence. To avoid this, comfort noise is generated matching the characteristic of background noise.
- *Discontinuous reception:* This is also a power-saving mechanism. While in idle mode, mobile station has to listen only to paging channel that does not consume significant power.
- *Handover:* When an MS is engaged to a call and moves it may go away from one BTS (BTS_1) and come closer to another BTS (BTS_2). As a result the signal strength received at MS from BTS_1 decreases whereas the signal strength from BTS_2 increases. When the signal strength from BTS_1 falls below a threshold, the control of the ongoing call goes from BTS_1 to BTS_2. The event of such switching of control of a call in progress from one BTS to the other is called handover. It will be elaborated in Section 3.3.4.

[11] BER is the ratio of the number of bits incorrectly received to the total number of bits transmitted during a specified time interval. For example, if 3 bits are erroneous out of total 10^5 bits, the BER is 3×10^{-5}.

The MM sublayer manages the problem arises out of mobility of a user. To keep track of a subscriber, the MM layer notes the location data of the user. It also takes care of the task of authentication and secured communication. Whenever a call is made for a mobile user, it is desirable to locate the user correctly. In one extreme, to solve this problem, the system has to page the whole network to locate a user. It requires a huge number of paging messages leading to wastage of scarce bandwidth. In the other extreme, the MS has to update the system on every move. This causes wastage of bandwidth due to a lot of obsolete update messages. The trade-off between these two extremes sometimes gives an optimal solution in some cases. GSM has implemented an optimal solution introducing a concept of location area. The location area is defined as a group of neighbouring cells. The critical issue is to select number of cells forming a location area. The solutions are mostly based on statistical data.

Whenever a user moves from one location area to the other, irrespective of the MS's state (idle/dedicated) it sends this update to the current VLR responsible for the MS. The VLR in turn informs the HLR of the MS about the update. While a call is in progress (MS is in dedicated state) and if the location area is changed the update is performed after the call is terminated. When an incoming call arrives for an MS, the current location area is paged to track the MS.

The CM sublayer manages circuit-oriented services such as call set-up, call maintenance and call termination. In addition to these services, it manages short messaging service.

The additional protocols used at A interface (Figure 3.6) are Signalling System No. 7 (SS7) and BSS Application Part (BSSAP). The SS7 is used between BSC & MSC, MSC & HLR/VLR and MSC & another MSC. This protocol transfers all management information among MSCs, HLR, VLRs, AC, EIR and OMC. This information exchange enables location update, handover, authentication, incoming call routing, etc. The BSSAP protocol is used for communication between an MSC and a BSC. RR messages are sent between BSC and MSC using the BSSAP. It manages the allocation of suitable radio resources to the MSs and tackles mobility management.

The additional protocol used at A_{bis} interface is BTS management (BTSM). The BTSM works between BTS and BSC and allows control of the radio equipment and radio frequency allocation to the BTS. However, the A_{bis} interface has not yet been standardized.

ISUP/R2: Referring Figure 3.5, ISUP (Integrated Services Digital Network User Part) is a part of the SS7 signalling protocol stack. ISUP/R2 is a variant of ISUP. It is used to establish call set-up involving PSTN. During a call set-up, using the ISUP/R2 signalling service, a switch sends the necessary call set-up information, e.g. the caller and called subscriber numbers, to the next switching point en-route.

3.3.3 Services and Features

GSM avails three classes of services—bearer services, teleservices and supplementary services. Other than these three, GSM provides a number of high data rate services. Two such services, namely HSCSD (high speed circuit switch data) and GPRS (General Packet Radio Service), are discussed elaborately in Section 6.2.

Bearer services: The services provide a reliable data transport connection and are called lower level services. Some of the bearer services are data, packet, etc. Such services can be (i) Transparent (T) or (ii) Non-Transparent (NT). The transparent services ensure constant bit rate with changing bit error probabilities whereas the non-transparent services activate an additional protocol between an MS and the MSC for resending blocks with observed errors.

Teleservices: The services like voice, SMS and Facsimile are the teleservices and use bearer services for transport. For example, the facsimile teleservice requires bearer service for error correction during facsimile transmission.

Supplementary services: Primarily supplementary services are of two types—call offering and call restriction. The common call offering service is Call Forwarding Unconditional (CFU). On the other hand, the example of call restriction service is Barring of All Outgoing Calls (BAOC).

3.3.4 Handover

The Handoff mechanism elaborated with respect to cellular mobile communication (Section 2.7) is referred to as Handover in GSM. Handover occurs due to switching of an ongoing call to a different channel. Primarily there are two categories of handover—(i) internal handover and (ii) external handover. The internal handover is initiated due to switching channels in the same cell or switching to a channel in different cell under the same BSC. Such handovers are internal to a BSC and involve only one BSC. The MSC is notified only on completion of the handover.

The external handover takes place in two cases—while switching channels from the cells under the control of two different BSCs but belonging to the same MSC, or while switching channels from the cells that are under the control of different MSCs.

Handover may be initiated by MSC or by an MS. The MS always scans broadcast control channels of up to 16 neighbouring cells and forms a list of the six best candidates based on the received signal strength for possible handover. This information is then transmitted to the current BTS at least once per second. The BSC and MSC use this information for making handover decision.

The received signal strength at MS is measured as low due to either physical interference or movement of MS to another cell. This makes difficult in taking handover decision. In GSM, although there is no recommendation

about when to perform handover, it is implemented by the following two techniques.

- If signal received by an MS from the servicing BTS degrades beyond some threshold, transmission power is to be increased. If power increase does not lead to improved signal quality, a handover is initiated. However, the increase in transmission power may cause interference with neighbouring cells.
- Each MS constantly measures signal strength from the servicing BTS as well as from neighbouring cell BTSs. Whenever the signal level of a neighbouring BTS is higher than the serving BTS, a handover is initiated. This technique avoids neighbouring cell interference but it is quite complicated.

3.3.5 GSM Channels

GSM uses two bands, each of 25 MHz (890–915 MHz and 935–960 MHz), for system use. The 890–915 MHz band is for uplink/reverse link and 935–960 MHz is for forward link. The available frequency bands are divided into 250 kHz wide channels, that is, the 25 MHz bandwidth can create 100 channels. If guard band of 125 kHz is considered on both sides of the designated spectrum, there can be $(25000 - 125 \times 2)/250 = 99$ available channels. In GSM, the bandwidth is divided among as many users as possible following a scheme that crossbreeds TDMA and FDMA. Each BTS is assigned one or more channels or carrier frequencies. Every channel is time shared among the eight users. The introduction of TDMA scheme enables partitioning of a channel in time, forming logical channels.

GSM views its channels as physical and logical channels. A physical channel corresponds to a segment of one or more radio frequency channels used to carry information. Such a segment is defined in terms of frequency, time, code, etc. depending on the multiple access technique used by the system. In GSM TDMA, the physical channels are determined by the time slot of a carrier frequency. On the other hand, logical channels are identified by the category of information carried within the physical channel. Each logical channel is mapped or multiplexed onto one or more physical channels.

Logical channels are used to handle either traffic data, signalling data or control data. The different categories of logical channels are shown in Figure 3.7. Accordingly logical channels are either traffic channel or control channel. The traffic channels and some of the control channels are dedicated. Rest of the control channels are referred to as common control channel. A dedicated channel provides a bi-directional point-to-point transmission link.

The common control channels are accessed by MSs both in idle and dedicated modes. An MS in idle mode accesses these channels to switch to dedicated mode whereas the MS in dedicated mode accesses common control channel to monitor surrounding BSs for handover information.

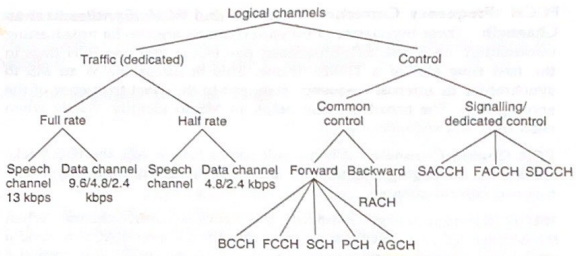


Figure 3.7 GSM logical channels.

The dedicated control channels are for maintenance of the call as well as for enabling a call set-up. These are utilized for managing handover, when the call is in progress, and finally for termination of the call. On the other hand, the traffic channels handle actual payload.

All the GSM logical channels are elaborated further below.

Traffic channels

Traffic channels carry digitally encoded user speech or data. It supports identical functions and formats on both the forward and reverse link. There are three types of traffic channels—TCH/F (full rate), TCH/H (half rate) and TCH/8 (one-eighth rate).

TCH/F: It transmits speech code of 13 kbps or three data-mode rates 12, 6 and 3.6 kbps.

TCH/H: It transmits speech code of 7 kbps or two data modes, 6 and 3.6 kbps.

TCH/8: It is used for low rate signalling channels, common channels and data channels.

Common control channels

A common control channel is unidirectional, that is, either from BTS to MS(s) or from MS to BTS. The tasks of all the common control channels are described below.

BCCH (Broadcast Control Channel): This is a forward control channel. BCCHs are for continuous information broadcasting about BTS identification, channel allocation, and list of adjacent BTSs. The channel allocation information helps an MS to know channel availability. On the other hand, with the help of the list of adjacent BTS information an MS can select one as its serving BTS. The BCCH is a point-to-multipoint channel from BSS to MSs.

FCCH (Frequency Correction Channel) and SCH (Synchronization Channel): These two classes of forward channels are also for broadcasting information. Each cell/BTS broadcasts one FCCH and one SCH flags in the first time slot of a TDMA frame. This broadcast helps an MS to synchronize its internal frequency standard to the exact frequency of the serving BTS. The broadcast also helps an MS to identify exactly when each time slot sequence begins.

PCH (Paging Channel): When a call comes for the MS, the BTS tracks the MS by the paging channel and alerts the MS of incoming call. It is a forward control channel.

RACH (Random Access Channel): It is a reverse control channel. When the MS desires to originate or set up a call, the MS uses RACH to send a request for accessing the network. In case of setting up a call in response to an alert received by PCH, the MS acknowledges the page received from the PCH by using RACH.

AGCH (Access Grant Channel): The AGCH, a forward control channel, is the final common control channel message sent by a BTS before the MS leaves all common control channels and uses dedicated channels. In response to RACH sent by an MS, the BTS uses AGCH to grant the request of the MS for accessing the network.

Dedicated control channels

All the channels in this group are bi-directional.

SDCCH (Stand-alone Dedicated Control Channel): In between granting access of network to the MS by BTS and assigning TCH to the MS, the BTS and MSC check the authentication of the MS. During this intermediate period when the MS waits for TCH, the BTS assigns SDCCH to the MS. If any information (e.g. authentication) to set up a call is sought from BTS during this period, the MS sends it over SDCCH.

SACCH (Slow Associated Control Channel): The SACCH is always associated with a traffic channel or a SDCCH. It provides a comparatively slow signalling connection. If the SACCH is associated with TCH, it is used for sending short message service (SMS).

FACCH (Fast Associated Control Channel): When a call is in progress, system information such as channel quality, power level, etc. that is necessary for call maintenance and handover are exchanged between BTS and MS over the FACCH. This channel works by stealing slots from a traffic channel.

Other than the three groups of channels mentioned above, Cell Broadcast Channel (CBCH) is an additional feature of a GSM system. It is a forward, point-to-multipoint logical channel. The CBCH is to support SMS broadcast service by means of which an MS can receive data/message broadcast, e.g. traffic, weather reports, etc. from the network service centre. It allows a limited short text as broadcast message.

3.3.6 Establishment of a GSM Call

There are three possible connections involving a GSM network—Land to Mobile (L_M), Mobile to Land (M_L) and Mobile to Mobile (M_M).

GSM-call originated by a land phone

In this case the call is originated by a Land phone and destined to an MS, i.e. L_M connection is to be considered. When a call originates from a Land phone, it is forwarded to the MSC through PSTN. The MSC checks from HLR/VLR whether the called MS is available in the network. If it is available, the MSC pages all the BSCs/BTSs under the control of the MSC to track the MS. From this point the BTSs take the responsibility for tracking the MS. A traffic channel is assigned for setting up the call after a number of message exchanges between BTS and MS. The activities of BTSs and MS to track the MS and (if the MS is tracked) to set up the call are summarized in the following steps.

- BTS broadcasts on paging channel (PCH) to track the MS
- MS
  - Detect the page
  - Reply the page and requests for accessing the network with a Random Access Channel (RACH)
- BTS
  - Grant the request over AGCH by assigning a channel SDCCH/SACCH to the MS
  - Request the MS over SDCCH to send data for authentication
- MS
  - Send authentication data on SDCCH
- BTS
  - Send call set-up and connection request over SDCCH
- MS
  - Acknowledge the request on SDCCH
- BTS
  - Assign TCH to the MS

GSM-call originated by a cell phone

In this case an MS originates the call whereas the call may be destined either to a land phone or to another MS, i.e. M_L and M_M connections are to be considered. When a call originates from an MS, it acquires a traffic channel to set up the call by exchanging messages between BTSs and MS. The following steps summarize the activities of the originating MS and BTSs.

- MS
  - Monitor the BCCH
  - Synchronize to a nearby BTS
  - Request to the BTS for accessing the network on RACH

- BTS
  - Respond the request on AGCH
  - Grant the request over AGCH by assigning a channel SDCCH/SACCH to the MS
  - Request the MS over SDCCH to send data for authentication
- MS
  - Send authentication data on SDCCH
- BTS
  - Send call set-up and connection request over SDCCH
- MS
  - Acknowledge the request on SDCCH
- BTS
  - Assign TCH to the MS

For the connection M_M, the originating MS gets permission of accessing the network and acquires traffic channel (TCH) by exchanging messages between BTS and MS as shown in 'GSM-call originated by cell phone'. On the other hand, the called MS gets permission to access the network and acquires TCH only after the MS is tracked. The message exchange between BTS and the called MS is performed as the steps shown in 'GSM-call originated by land phone'.

3.3.7 Channel Usage during GSM Call

In case of L_M and M_M connections, that is when called unit is an MS, all the tasks shown in Figure 3.8(a), (b) and (c) are done for providing channel to the called MS to set up and release the call. However, for M_L and M_M connections, that is when call is originated by an MS, the tasks shown in Figure 3.8(a) and (c) are performed for providing channel to the originating MS to set up and release the call. Therefore, for M_M connection first the tasks (a) and (c) are performed for giving channel to the originating MS, and then all the tasks (a), (b) and (c) are performed for providing channel to the called MS.

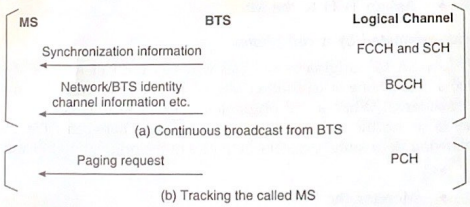


Figure 3.8 *Contd.*

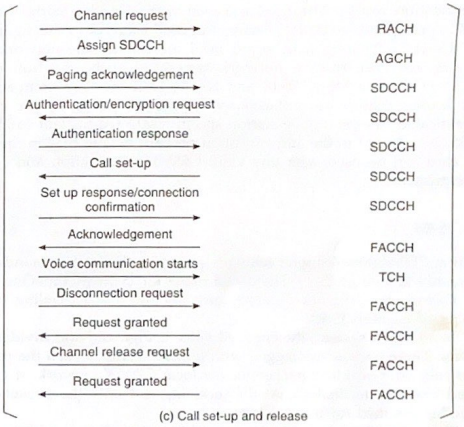


Figure 3.8 Usage of GSM channels.

3.3.8 User Validation in GSM

In GSM, subscriber identity module (SIM) of an MS contains international mobile subscription identity (IMSI), authentication key K_i, A_3 algorithm and A_8 ciphering key generating algorithm. The authentication process of GSM-based systems check the validity of SIM. It is analogous to a challenge-and-response process.

The GSM network sends a 128-bit random number to an MS as the challenge. The MS feeds the random number (challenge) and K_i to the A_3 algorithm to generate a signed response (SRES). The SRES in turn is sent back to the network. The network compares the received SRES with the SRES supplied by the authentication centre (AC). If there is a match, the subscriber is recognized as valid. Upon confirming the user validity, the A_3 algorithm is run at the both ends of the air interface to generate a ciphering key (K_c) considering the random number and K_i as the inputs. The K_c is used as input to the A_8 algorithm for encryption and decryption of data.

In GSM-1900/AMPS, a dual mode MS, the user validation process is a key-based authentication process similar to that as discussed in Section 2.11.2. Here, a mobile station intending to make a call runs an algorithm to generate a data called MS-AUTHR (mobile station generated

authentication result). The IMSI is keyed with a hidden authentication key K_i (corresponds to AMPS A-key) and used as input to the algorithm. The MS-AUTHR along with keyed IMSI are sent to the network. The network also generates a network generated authentication result (NT-AUTHR). The MS-AUTHR and NT-AUTHR are then compared. If underscores should be replaced by hyphen they match, the call is authenticated. As the mobile station specific IMEI (equivalent to ESN of AMPS) is not used in the authentication process, a valid SIM in the form of a card can be used with any valid GSM mobile station and can be authenticated.

3.4 IS-95

The first CDMA-based digital cellular standard is the Interim Standard 95 commonly known as IS-95. The brand name for IS-95 is cdmaOne. As it uses CDMA, the network capacity does not put a strict ceiling on the number of users in IS-95.

In a CDMA system, the encoded voice is digitized and divided into packets. These packets are tagged with unique "codes". Then the packets are mixed with all other packets in the local CDMA network. It is then routed towards destination. At the receiving end only the packets with the codes destined for it are accepted.

The users share a common channel for transmission within a cell. Users in adjacent cells also use the same radio channel. In other words, the frequency spectrum is reused. The spreading factor used in IS-95 is 128 with the maximum user data rate 9.6 kbps. Forward and reverse links use different spreading techniques. As discussed in Section 3.2.3, optionally IS-95 employs two-phase or multiple spreading. One phase provides mutual orthogonality among all users in one cell and the other phase provides mutual orthogonality among the users in different cells. Rake receivers[12] are used at both the base station and mobile station to resolve and to combine multi-path components.

3.4.1 System Architecture

Figure 3.9 shows the IS-95 system architecture including interfaces. This section describes the tasks performed by each module of the system.

Inter-Working Function (IWF): It is the interface (L of Figure 3.9) between a wireless system and the telephone network. IWF converts data, transmitted over the air interface, to a format recognized as well as carried by the public telephone network, e.g. PSTN. It also synchronizes transfer between

[12] In DSSS, receivers face a problem during reconstruction of the original data due to multi-path propagation. Signal reaches to a receiver from multiple paths having different delays and path loss. Rake receivers apply a mechanism to solve this problem and reconstruct the original data.

a circuit-switched network and the packet-switched network, and that effectively enables an MSC to communicate with other networks.

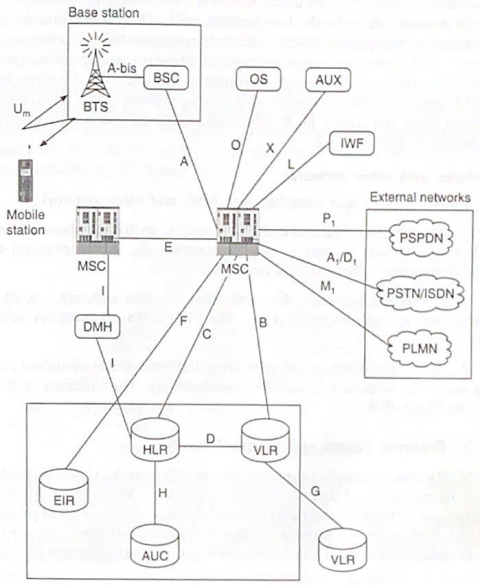


Figure 3.9 IS-95 architecture.

Operation System (OS): The OS is responsible for overall management of the wireless network. Its interface with MSC is denoted as 'O'.

Auxiliary (AUX) equipment: An MSC is connected with a number of auxiliary equipments such as signal transfer point (STP), short message service centres (SMSCs) and voice mailbox system. The interface is denoted as 'X' in Figure 3.9.

Data Message Handler (DMH): The responsibility of DMH is to collect billing data. DMH is connected with MSC by an interface 'I'. It provides: (i) a standard for call detail and billing record format, (ii) procedures and messages required to perform such record transmission between systems.

HLR/VLR/AUC: HLR (Home Location Register), VLR (Visiting Location Register) and AUC (Authentication Centre) are involved in mobility management, handover, and user validation. The tasks of these components in GSM system are described in Section 3.3.1. The HLR contains static database comprising subscribers' information such as ESN (electronic serial number), IMSI (international mobile station identity), user profiles, present location of a user, etc. to manage subscribers' mobility. On the other hand, the VLR keeps dynamic database comprising visiting users' ESN, profiles collected from the user's HLR. The AUC takes care of user validation, elaborated in Section 3.4.4.

Interfaces with other networks

The following interfaces exist between MSC and other networks.

A_1/D_1: The MSC is connected with PSTN (public switch telephone network) by analog interface A_1 and by digital interface D_1, it is connected with ISDN (integrated switch digital network).

M_1: The interface between MSC and other wireless network, i.e. PLMN (public land mobile network) is M_1. The PLMN may be another cellular network.

P_1: An MSC can communicate with other PSPDN (packet-switched public data network) such as X.25 and IP-based network. The interface is shown as P_1 in Figure 3.9.

3.4.2 Protocol Layers and Channels in IS-95

The IS-95 standard does not explicitly mention the functions of each protocol layer. However, the functions of each layer very much exist. In the air interface, i.e. between a mobile station and the base station, a set of protocols is used. It describes a three-layer stack—(i) physical layer, (ii) medium access control (MAC) and link access control (LAC) sublayers and (iii) upper layer.

(i) Physical layer

The IS-95 standard defines the transmission of signals in physical layer in both the forward (downlink) and reverse (uplink) directions. Every IS-95 service provider receives 12.5 MHz spectrum. The 10% of available cellular spectrum, i.e. 1.25 MHz is occupied by each channel. Unlike other cellular standards, the user data rate (but not the channel chip rate) changes in real time depending on the voice activity and requirements of the network.

Forward channels

In the forward direction, radio signals are transmitted by BTSs. Every BTS is synchronized with a GPS receiver. All forward transmissions are BPSK[13] (binary phase shift keying) with a chip rate of 1,228,000/s. Multiple spreading

[13] The simplest form of phase shift keying considering two phases.

is used for this transmission. Each signal is spread with a walsh code of length 64 and a pseudo-random noise code (PN code) of length 2^{15} causing a PN roll-over period[14] of 80/3 ms. The walsh codes differentiate transmissions within a cell. The PN codes are used to isolate cells (BTSs) that are using the same frequencies. The same PN sequence is used in all BTSs and the offset for each BTS is unique.

Forward broadcast channel: A forward channel consists of pilot channel, synchronization channel, maximum seven paging channels and at most sixty-three forward traffic channels. The pilot, synchronization and paging channels are considered as the broadcast channel. The details of IS-95 channels are shown in Figure 3.10.

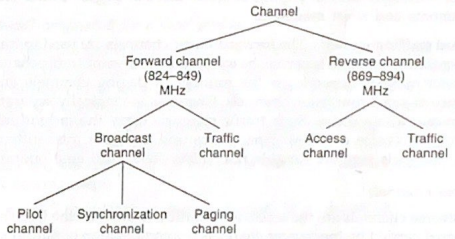


Figure 3.10 IS-95 Channel structure.

Pilot channel. The *pilot channel* consists of an unmodulated PN sequences spread with walsh code 0. The power control is not required in the pilot channel. It does not carry any data. Each BTS in the network is assigned a PN offset in steps of 64 chips. Each pilot transmits the same spreading sequence at different time offset.

Each MS continuously monitor pilot channels. With the help of broadcast information sent over pilot channel, an MS distinguishes different BTSs comparing signal strength received by the MS. The monitoring of pilot channel and measurement of signal strength received from different BTSs by an MS help to take decision of handoff.

Synchronization channel. A *synchronization (sync) channel* provides system parameters to an MS that are required to synchronize with the network and obtain a paging channel. The channel continually broadcasts sync messages for the MSs. The messages contain information about the network including its identity, version of radio interface being supported, PN offset used by BTS, etc. It uses walsh 32 for spreading and operates at 1200 bps. It also helps the MSs for acquiring a paging channel.

[14] The period with which the sequence reproduces.

Once an MS finds a strong pilot channel, it listens to the sync channel and decodes the sync channel message to build up a highly accurate synchronization with the system time. As the sync channel message contains network-ID, the MS can know whether it is currently in roaming.

Paging channel. The *paging channel* carries either control information for call set-up or paging messages, when an incoming call from BTS to MS is to be served by them. It does not require any power control. There are three possible rates used in the paging channel—9600, 4800 and 2400 bps. All the rates are encoded to 19200 symbols per second. The messages on paging channel convey detailed network parameters and also carry higher-priority messages dedicated to setting up a call to and from the MSs. Typical messages on the paging channel include pages, traffic channel assignments and short messages.

Forward traffic channel: The forward traffic channels are used to transmit voice or data to an MS. There may be maximum 63 forward traffic channels—the exact number depends on the number of paging channels and the presence of synchronization channels. Channels are logically separated by the unique walsh codes. Such traffic channels carry the individual user information (voice or data). Since voice and data are intermittent, the traffic channels support variable rate (1200, 4800, 9600 bps) operation.

Reverse channels

The reverse channels are the access and traffic channels. At the BTS receiver the signal carried by the reverse channels is a combination of output signals from all the MSs within the BTS's coverage area. The reverse channel allows up to 62 traffic channels and 32 access channels. However, the number of channels in use at any point of time may be considerably lower.

Reverse traffic channel. Reverse channels use OQPSK (offset quadrature phase shift keying) for power efficiency. All data transmitted on the reverse channel are convolutionally encoded, block interleaved, modulated by a 64-ary orthogonal modulation and spread prior to transmission.

Access channel. An access channel allows an MS to communicate with the system when it needs to initiate an action such as registration, call origination or it needs to respond to the messages received on a paging channel. Since multiple MSs may attempt access at the same time, the access channel utilises suitable protocol for contention management. The data rate for the access channel is 4800 b/s.

(ii) MAC and LAC sublayers

The MAC sublayer provides a control function to manage resources supplied by the physical layer and coordinates their usage by LAC sublayer. Once a call is established, an MS uses traffic channel that allows voice or data bits to be multiplexed with signalling message fragments. The signalling

message fragments are assembled in the LAC. The LAC in turn sends complete signalling messages to the upper layer. The MAC provides multiplexing and QoS control. The QoS is implemented by prioritizing requests and resolving conflict messages.

(iii) Upper layer

The overall control of the system is taken care of by the upper layer. Both the voice and data messages pass through the upper layer.

3.4.3 Establishment of a IS-95 Call

Like GSM, there are three possible connections involving a IS-95 network—Land to Mobile (L_M), Mobile to Land (M_L) and Mobile to Mobile (M_M).

IS-95-call originated by a land phone

In this case the call is originated by a Land phone and destined to an MS that is L_M connection is to be considered. Here, the call is forwarded to BTSs through MSC in the same way as described in Section 3.3.6. Once the call is forwarded to the BTSs, the BTSs take the responsibility for tracking the MS. A traffic channel is assigned for setting up the call after exchanging messages between BTS and MS. The activities of BTSs and MS to track the MS and set up (if the MS is tracked) a call are summed up in the following steps.

- BTS broadcast on paging channel to track the MS
- MS
  - Detect the page
  - Acknowledge on access channel
- BTS
  - Configure the transmitter and receiver by walsh code and PN code respectively
  - Send this information on paging channel
- MS
  - Receive paging channel
  - Configures transmitter and receiver by PN code and walsh code respectively
  - Send authentication information on access channel
- BTS
  - Authenticate the MS
  - Send an alert message for ringtone
- MS
  - Acknowledge the alert by ringing
- BTS
  - Assign traffic channel to the MS

IS-95-call originated by a cell phone

In this case an MS originates the call whereas the call may be destined to either a Land phone or to another MS that is M_L and M_M connections

are to be considered. When a call originates from an MS, it acquires a traffic channel to set up the call by exchanging messages between BTSs and MS. The following steps sum up the activities of the originating MS and BTSs.

- MS
  - Monitor the pilot and sync channel
  - Synchronize to a network BTS
  - Seek network access permission on access channel
  - Monitor paging channel for BTS response
- BTS
  - Grant the access request on paging channel
- MS
  - Send authentication information on access channel
- BTS
  - Authenticate the MS
  - Assign traffic channel to the MS

Like GSM, in case of M_M connection, the originating MS gets permission of accessing the network and acquires traffic channel by exchanging messages between BTS and originating MS as in 'IS-95-call originated by cell phone'. On the other hand, the destination MS gets permission to access the network and acquires traffic channel only after the MS is tracked. The message exchange between BTS and destination MS is performed following 'IS-95-call originated by land phone'.

3.4.4 User Validation in IS-95

The user validation process is similar to the process used for GSM (discussed in Section 3.3.8). Instead of A_3 algorithm in GSM, the CAVE (cellular authentication and voice encryption) algorithm is stored in an MS. The network and the MS share a secret key referred as SSD (shared secret data) consisting of two parts: SSD-A and SSD-B, each of 64 bits. A BTS broadcasts time to time a 32-bit random number as the so-called challenge. Each MS feeds the received challenge, SSD-A, ESN (electronic serial number) and MIN (mobile identification number) to the CAVE algorithm to obtain a SRES (signed response). The SRES is then sent to the network's authentication centre (AUC). If there is a match between SRES from MS and the SRES computed at AUC, the MS is authenticated. Once authentication is over, the network sends encryption key and voice privacy mask[15] to the MS. The encryption key and voice privacy mask are used for encryption of signalling message and voice respectively.

[15] In IS-95, each MS is provided a unique long code as mask based on its ESN. This mask is known as voice privacy mask that is applied to the voice data to provide privacy at both downlink and uplink data transfer over air interface.

BIBLIOGRAPHY

Abramson, N., *Multiple Access Communications: Foundations for Emerging Technologies*, IEEE Press, New Jersey, 1993.

Bernard J., and T. Mallinder, "An Overview of the GSM System", *Proceedings, Digital Cellular Radio Conference*, Hagen, FRG, 1988.

Black, U., *Second-generation Mobile and Wireless Networks*, Upper Saddle River, New Jersey, Prentice-Hall, New Jersey, 1999.

Cellular System, IS-95, "Dual-Mode Mobile Station-Base Station Wideband Spread Spectrum Compatibility Standard", PN 3118, EIA, Engineering Department, CDMA System, 1992.

Dechaux, C., and R. Scheller, "What are GSM and Dect?", *Electrical Communication*, pp. 118–127, 2nd quarter, 1993.

Dinan, E., and B. Jabbari, "Spreading Codes for Direct Sequence CDMA and Wideband CDMA Cellular Networks", *IEEE Communications Magazine*, 1998.

ETSI "European Digital Cellular Telecommunications System (Phase 2): General Description of a GSM Public Land Mobile Network", GSM 01–12, 1993.

Falconer, D.D., F. Adachi and B. Gudmundson, "Time division multiple access methods for wireless personal communications", *IEEE Communications Magazine*, Vol. 33, No.1, pp. 50–57, 1995.

Garg, V., and J. Wilkes, *Principles and Applications of GSM*, Englewood Cliffs, Prentice-Hall, 1999.

Gilhousen, et al., "On the capacity of cellular CDMA System", *IEEE Transactions on Vehicular Technology*, Vol. 40, No. 2, pp. 303–311, 1991.

Gilhousen, K.S., I.M. Jacobs, R. Padovani, A.J. Viterbi, L.A. Weaver and C.E. Wheatley, "On the capacity of a cellular CDMA systems", *IEEE Transactions on Vehicular Technology*, Vol. 40, No. 2, 1991.

Gudmundson, B., J. Skold and J.K. Ugland, "A Comparison of CDMA and TDMA systems", *Proceedings, IEEE Vehicular Technology Conference*, Vol. 2, pp. 732–735, 1992.

Hodges, M.R.L., "The GSM Radio Interface", *British Telecom Technological Journal*, Vol. 8, No. 1, pp. 31–43, 1990.

Lee, W.C.Y., "Overview of Cellular CDMA", *IEEE Transaction on Vehicular Technology*, Vol. 40, No. 2, 1991.

Lee, William C.Y., *Mobile Cellular Telecommunications Analog and Digital Systems*, New Delhi, McGraw-Hill, Inc., International Editions, 1995.

Modarressi A.R., and R.A. Skoog, "Overview of signalling system no. 7 and its role in the evolving information age network", *Proceedings IEEE*, New York, Vol. 80, No. 4, pp. 590–606, 1992.

Mouly, M., and M.B. Pantel, "Current Evolution of GSM systems", *Personal Communications Magazine*, Vol. 2, No. 5, 1995.

Murota K., and K. Hirade, "GMSK modulation for digital mobile radio telephone", *IEEE Transactions on Communications*, Vol. 29, No. 7, pp. 1044–1050, 1981.

Noerpel, A., and Y.B. Lin, "Handover management for a PCS network", *IEEE Personal Communications*, Vol. 4, No. 6, pp. 18–26, 1997.

Pandya, Raj, *Mobile and Personal Communication Systems and Services*, New Delhi, Prentice-Hall of India, 2004.

Peterson, R.L., R. Ziemer and D. Borth, *Introduction to Spread Spectrum Communications*, Englewood Cliffs, New Jersey, Prentice-Hall, 1995.

Pickholtz, R.L., L.B. Milstein and D. Schilling, "Spread Spectrum for Mobile Communications", *IEEE Transactions on Vehicular Technology*, Vol. 40, No. 2, pp. 313–322, 1991.

Pollini, G.P., "Trend in handover design", *IEEE Communications Magazine*, Vol. 34, No.2, pp. 82–93, 1996.

Prasad, R., and T. Ojanpera, "An overview of CDMA Evolution: Toward wideband CDMA", *IEEE Communications Surveys*, 4th quarter, 1998. website: http://www.comsoc.org

Rahnema, M., "Overview of the GSM System and Protocol Architecture", *IEEE Communications Magazine*, 1993.

Raith, K., and J. Uddenfeldt, "Capacity of Digital Cellular TDMA Systems", *IEEE Transactions on Vehicular Technology*, Vol. 40, No. 2, pp. 323–331, 1991.

Rappaport, Theodore S., *Wireless Communications Principles and Practice*, Prentice-Hall PTR, New Jersey, 1999.

Ross, A.H.M., and K.S. Gilhousen, " CDMA Technology and the IS-95 North American Standard", *Mobile Communication Handbook*, J.D. Gibson (Editor), CRC & IEEE Press, 1996.

Salmasi A., and K.S.Gilhousen, "On the System Design Aspects of Code Division Multiple Access (CDMA) applied to Digital Cellular and Personal Communications Networks", *IEEE Vehicular Technology Conference*, St. Louis, USA, pp. 57–62, 1991.

Schiller, Jochen H., *Mobile Communications*, New Delhi, Pearson Education, 2007.

Scholtz, R.A., "The spread spectrum concept", *IEEE Transactions on Communications*, Vol. 25, pp. 748–755, 1977.

Simon, M.K., J.K. Omura, R.A. Scholtz and B.K. Levitt (Editors), *Code-Division Multiple Access, Spread Spectrum Communications Handbook*, New York, McGraw-Hill, 1996.

Tantaratana, S., and K. Ahmed (Editors), "Wireless Applications of Spread Spectrum Systems: Selected Readings", IEEE Press, 1998.

Viterbi, A., *CDMA: Principles of Spread Spectrum Communication*, Redwood City, CA, Addison-Wesley Longman, 1995.

REVIEW QUESTIONS

1. What is the purpose of multiple access technique?
2. What is guard band? Why is it used? Explain it citing an example.
3. What is the principle of TDMA system? How can it accommodate more users using the same spectrum as FDMA?
4. What is the principle of CDMA? How a particular user's data is decoded at the base station? Explain it citing an example.
5. How does system capacity increase to a large extent in CDMA system compared to TDMA and FDMA system?
6. Explain the difference between autocorrelation and cross correlation.
7. What is spread spectrum technique? List the benefits of spread spectrum. Describe DSSS CDMA-based modulation/demodulation technique with the help of a block diagram.
8. How does the near-far effect influence CDMA systems? What is the countermeasure?
9. What are the tasks performed by HLR, VLR, AC, EIR in GSM?
10. Write the full forms of the following and state the task of each of them.
    SIM
    IMSI
    TMSI
    IMEI
11. Represent GSM architecture diagrammatically. Describe the functions of each of the modules.
12. Who can initiate handover? What are the techniques followed to implement handover in GSM?
13. What is the basic difference between the tasks handled by traffic channels and the tasks handled by control channels?
14. Name the bi-directional control channels in GSM. State their contributions in setting up of a GSM call.
15. In which category the following GSM channels belong?
    BCCH
    FCCH
    PCH
    RACH
    AGCH
    Describe their responsibilities to establish and release a GSM call.
16. Write step by step message exchange that takes place among the elements of GSM network to establish connection between the two GSM mobile stations.
17. Diagrammatically represent the steps of operation for validating a user in GSM.

18. (a) What is the full name of IS-95? What multiple access technique does it use?
    (b) What is multiple spreading? Why is it used in IS-95?
19. Considering IS-95 architecture, name the module that is used as interface between the IS-95 network and telephone network (PSTN). What are the tasks performed by this module?
20. (a) Which module in IS-95 is responsible for collecting billing data?
    (b) List the interfaces exist between IS-95 and other networks such as PSTN, PSPDN, etc.
21. What is the role of synchronization channel in IS-95?
22. Explain how a user is validated in IS-95.

4

WIRELESS METROPOLITAN AREA NETWORK (WIRELESS LOCAL LOOP)

Mobile Computing deals with computing on move. Users with a portable device can access a network or other users in the network. So it is implied that the users on move can communicate as well. It is also implied that the portable communicating device communicates without a wire, i.e. wireless. Therefore, without the discussion of mobile services/technology based on various networks, a book on mobile computing is incomplete. As discussed in Chapter 1, according to the coverage area the networks are categorized into three types—WAN, MAN and LAN. Cellular network, as an example of WAN, has been discussed previously. Wireless Local Loop (WLL) is discussed here as an example of MAN. Instead of contiguous large area coverage as in WAN, MAN coverage is restricted to a comparatively smaller area, e.g. a metropolitan city. WLL is a technology for providing services in telecommunication domain for subscribers in a cluster of locality, e.g. a city.

4.1 HISTORY

Traditionally, the communication between a subscriber (telephone) and the local telephone exchange, called local loop, is realized with a pair of copper cables. At the initial stage of development, the local loop used to be as long as 8 to 9 kms. The copper wire connection is very costly and it is over 80 per cent of the total cost of a new connection. In the 1980s the newly introduced technology enabled shortening of local loop to about 3 kms. The introduction of remote line unit (RLU) or remote switching unit (RSU) multiplexing facility permitted several subscribers to share a common transmission medium and thereby, reduced the communication cost. However, all these cannot deny the basic requirement of a distinct line for each subscriber even if it is of a very short length.

Since the 1980s, the demand for communication services is increasing explosively. The services include the basic telephone service as well as the high-speed data and multimedia services. The initial attempt to cope up with the requirements was based on the concept of local loop. The local loops in wired environment properly responded to these increasing demands through the use of digital technologies such as ISDN (Integrated Services Digital Network) and DSL (Digital Subscriber Line). However, in the mid 1990s, the universally accepted codes and protocols were enforced to govern the transmission of signals from subscribers to exchanges. The standardized signal processing equipments were now available at both the ends, and this allowed any kind of transmission medium, including radio frequency spectrum to replace the copper wire. On the other hand, with the advent of Internet, the demand for network complexity has been growing rapidly. The wireless in local loop (WLL) has thus evolved as the natural choice.

4.2 LIMITATIONS OF TRADITIONAL TELEPHONE NETWORK

Nowadays small offices and homes have Internet access (Figure 4.1) based on the public switch telephone network (PSTN). It simply requires the subscriber to take a telephone line, a modem and a computer to set up a connection. The subscriber is to dial an Internet Service Provider (ISP). The ISP owns a number of telephone lines and modems and connects a user to a router. The router finally allows access to the Internet. The dial-up connection is easy to install. However, such an Internet system realized with traditional telephone network has the following limitations:

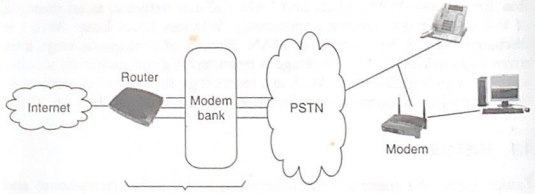


Figure 4.1 Internet access with PSTN.

(i) For every subscriber, the telephone network is designed to handle a limited number of traffic adequate for telephony. On the other hand, Internet sessions desired by a subscriber are usually of long duration. Therefore, for a reasonable number of Internet subscribers, the telecom network cannot be suitable in terms of traffic load handling capacity, and that results in severe congestion.

(ii) The analog modem to modem link between the subscriber and the ISP in such a set-up is unreliable. One does get a 33.6 kbps connectivity sometimes, but connection could go down to 9.6 kbps and even 4.8 kbps at times. Connection may be dropped as well.

(iii) The volume of services provided by a dial-up Internet connection system is limited by the number of available telephone lines and modems in ISP. An ISP with N telephone lines, N modems and an N port router can serve at most N subscribers at a time.

An alternative solution is to implement totally shared packet-switched Internet access. In Internet access, packets are transmitted in bursts. During an Internet session, the communication between a subscriber and the service provider remains idle for most of the time. A circuit-switched connection on telephone network cannot utilize the idle slots; it rather occupies resources throughout the session. This causes congestion in the network. In circuit-switched connection a number of new technologies such as ISDN and xDSL are introduced in an attempt to tune the wired technologies to respond with the need of reliable, high-speed access by users connected in fixed lines. However, the data access has been made more realistic providing packet-switched access.

As the local loop is the separate physical line to each subscriber, packet access on such non-sharable resource cannot have additional advantages. More than one subscriber is not allowed to use this resource. Therefore, to get the advantage of packet-switched access, Internet data is separated at a point (P) closest to the subscriber where data from multiple subscribers can be multiplexed. The connection from a subscriber to P is made by shared medium access that is WLL.

4.3 APPLICATION DOMAIN

The WLL is quicker, less expensive and easier to install compared to a conventional landline system. This creates enormous changes in

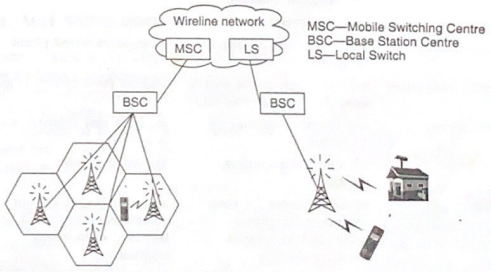


Figure 4.2 Cellular vs WLL system.

communication services in developing countries. It is very much effective where installation of landline system is not even feasible. However, for a developed country, WLL may be used in cordless phone as an extension of house/office telephone or private branch exchange (PBX). This adds convenience in the use of telephonic system (Refer to Figure 4.2).

The WLL can be installed and dismantled quickly (in weeks) and easily, and is ideal also for a temporary solution, instead of waiting for the deployment of copper or fibre cables. Different WLL solutions are introduced to satisfy different needs in terms of service, coverage, frequency spectrum, subscribers' density, etc.

4.3.1 Mobile Cellular System vs WLL System

Although there are similarities between a WLL system and the mobile cellular system as both of these support mobile telephony, the basic difference lies in the planning and design considerations. A WLL system is considered for low-mobility users and designed either as large fixed radio access networks or as the clusters of micro cellular networks. The voice quality in a WLL is comparable with that of a landline telephone system. On the contrary, the cellular system aims at a large coverage area around a base station and can accommodate subscribers with high mobility. However, the voice quality is sacrificed.

The following table notes the expectations from a mobile cellular system and the WLL:

Parameter	*Expectations from*	
	Mobile cellular system	*WLL system*
Frequency bands	regulated dedicated bands	no dedicated bands
Radio interfaces	conform to regional and international standard	not standardized
Coverage	wide/universal	limited
Voice quality	a modest voice quality is acceptable	as good as wired phone
Traffic/subscriber	not very high as the user is unlikely to make long calls	high as voice quality is good
Data service	low bit-rate data may be acceptable	medium rate Internet access is a must
Mobility	full mobility including roaming	limited mobility
Cost	air-time charges for some services is acceptable; higher cost to support mobility/roaming	air-time charges are not acceptable; cost should be less than even wired telephone

4.3.2 Merits of Adopting WLL

The WLL systems can be deployed in weeks, and hence it is far easier to implement than the systems designed with copper wire. Faster deployment of WLL leads to realization of revenues sooner as well as fast recovery of the deployment cost. Further, the deployment of WLL involves considerably less costly construction than the laying of copper lines in conventional telephone system. The operations and maintenance in WLL are also easy. The average instances of maintenance per subscriber per year are 3 to 4 times shorter than their wireline implementations.

In WLL, the connection time to accommodate a new subscriber is much lower than that in a wireline or cellular system. This gives satisfaction to a subscriber and significantly reduces the overall cost for each customer.

Moreover, the use of advanced digital radio technology enables the WLL to provide variety of high bandwidth data, multimedia and voice services.

In a wireline telephone network, any form of extension targeting new subscribers in an area/domain demands considerable additional investments. However, in WLL once the WLL infrastructure-network of base stations and interface to the telephone network are in place, adding a new connection (subscriber) involves almost no additional investment. Further, most of the WLL systems are designed to be modular and scalable that allow a network operator to keep pace with the incremental demand, simultaneously ensuring a minimum loss due to investment associated with the underutilized network resources. So a WLL system is very much flexible to meet even the uncertain rates of growth.

In a word, apart from fast deployment and providing high bandwidth services, various costs associated with the WLL such as construction, maintenance and network extension costs are comparatively lower than wireline and cellular systems.

4.4 WLL TECHNOLOGY

There is no specific band that WLL systems occupy or are deployed in. The systems can either operate in a dedicated, protected spectrum or in an unlicensed spectrum. Some of the services that fall within the definition of WLL include cordless phone systems and fixed cellular systems as well as a variety of proprietary systems. The WLL technology is commonly based on: (i) cellular mobile radio standards, (ii) cordless mobile radio standards, (iii) proprietary WLL technologies.

The low price of cellular transmission equipments due to its production in large scale makes the cellular technology attractive for application in wireless local loops. The commonly used cellular telephony standards in designing WLL are AMPS, GSM and IS-95. The introduction of cellular technology in WLL simplifies the system design in comparison to the cellular telephony, as mobility is not considered in WLL normally. As

mobility is not considered, handover and roaming do not occur. So the design of system's network layer is considerably simplified. However, the quality of the offered services in WLL strongly depends on the applied technology. The cellular systems used to implement the WLL are effective for realization of simple telephone services in large low-populated areas.

The example of a WLL system that uses the radio interface as cellular mobile system is STRAEX WLL system of South Korea. On the other hand, AIRLOOP (USA) uses proprietary radio transmission technologies. There are unused frequency bands available above 25 GHz. The frequency range used in WLL is 10 to 300 GHz. These wide channel bandwidths can be used providing high data rates with small size transceivers. However, it faces free space loss and multi-path loss. Further, beyond 10 GHz, attenuation due to rainfall and atmospheric absorption cannot be neglected. These characteristics limit the range of service area covered by a WLL cell. Usually, the cell radius is about a few km.

4.4.1 WLL System Architecture

Figure 4.3 describes the architecture of a WLL system. The major components of such a system are the End-user Terminal, Fixed Subscriber Unit (FSU), Base Station System (BSS), Network Management Unit (NMU), Operation Administration Maintenance (OAM) and interfaces.

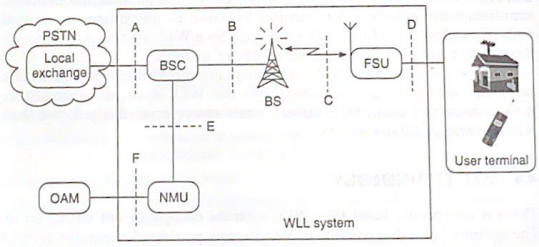


Figure 4.3 WLL system architecture.

End-user terminal: The radio terminals in the system, that is base station and user terminal, consists of the basic components such as indoor unit (IDU), outdoor unit (ODU) and antenna. The IDU and ODU are dedicated for transmission and reception of the outgoing and incoming user data. Like any other radio communication system, the transmitted/received signal passes through different phases (Section 1.2.4) from source coder/demodulator to modulator/decoder. In transmission, the indoor unit converts an analog signal (e.g. voice) to digital, mitigate the interference and errors (if any) and transforms to a suitable form for transmission by the outdoor unit and the antenna. In reception, a reverse process of IDU/ODU takes place.

During transmission, the IDU deals with signal conversion, interference mitigation, and transferring the signal to ODU. On the other hand, the task of ODU, during transmission, is to modulate the frequency signal received from the IDU to a radio frequency signal at the appropriate frequency for transmission. Similarly, during reception, an ODU demodulates the received radio frequency signal to a suitable frequency acceptable to the IDU. The IDUs and ODUs are usually connected via coaxial cables. The BSs antennas are normally sectoral antennas and at the end-user terminal it is high-gain directional antennas. Such arrangement minimizes interferences as well as improves network coverage area and capacity. However, the directional antennas at the end-user terminals limit the portability of the terminals and, therefore, may not be effective for WLL system supporting mobility.

Fixed Subscriber Unit (FSU): The FSU is the interface between the user terminals and WLL network. It is designed to support different types of end-user terminals (Computers, Facsimiles, Telephones, etc.) so that these can communicate with the BS through radio signals. Therefore, varieties of FSU implementations are required for WLL. In one implementation, FSU is a separate centralized unit catering many subscribers located usually within a multi-storeyed building, local business centre, etc. In some recent products, handset with in-built FSU is also available. The difference between a handset with integrated FSU and the handset used for mobile cellular communications is that the former does not include the mobility management functionality.

Base Station System (BSS): A base station system is the combination of base station (BS) and base station controller (BSC). A BS provides radio service to the end-user terminals within the cell area covered by the BS. It configures a point (BS) to multipoint (terminals) radio communication system. On the other hand, the BSC is an interface between the WLL system and a telecommunication network. There are two approaches in designing the BSC—hierarchical and distributed. In the hierarchical design, a good number of BSs is served by the single BSC. The BSC in turn connects backbone network, e.g. PSTN. On the contrary, in a distributed architecture, each BS has direct communication link with the PSTN.

NMU (Network Management Unit): The NMU as shown in Figure 4.3 performs data management related to the system configuration based on customer, system, and radio link parameters. It also monitors the local loop status to reconfigure the network and its traffic flows for avoiding interference and failures.

OAM (Operation Administration Maintenance): The OAM module is used by the network operator to maintain and configure the network. The maintenance tools commonly available in an OAM performs: (i) remote configuration of nodes (ii) subscriber management—addition and removal of subscribers, temporarily disabling subscribers, etc. (iii) performance

management controlling traffic in the network (iv) billing for the subscribers, and (v) switching functions. The protocol commonly used in an OAM is the simple network management protocol (SNMP[1]).

Both the NMU and OAM units are for system management. In some implementations/products, NMU and OAM are integrated and in other implementations they are treated separately.

Interfaces: The A, B, C, D, E and F in Figure 4.3 indicate the main interfaces in WLL system architecture. The devices BSC and BS are normally manufactured and supplied by the same vendor. It is also true for user terminal and FSU, OAM and NMU. The corresponding interfaces are considered as proprietary interfaces. Therefore, all the interfaces, except A and C, are proprietary in nature and are provided by the device manufacturer.

The FSU at end-user side communicates base station via 'C' radio interface. The tasks of this interface are source coding/decoding, channel coding/decoding, modulation/demodulation (Section 1.2.4) for successful transmission/reception of signal over radio. In spite of the fact that radio systems both at end-user side and at BS perform the said tasks, parameters such as antenna type, transmitting power level, etc. are different in both the radio systems. The radio system at BS has to bear some additional responsibilities such as radio resource allocation/deallocation, call establishment/release and mobility management (if required). The WLL network is connected with fixed telephone network (PSTN) via 'A' interface between BSC and the PSTN. The main function performed at this interface is conversion of source code used in fixed network into the source code suitable for wireless network and vice versa.

The advantage of WLL systems based on cellular or cordless radio standards over proprietary radio transmission technologies is that the former is the by-product of cellular or cordless system. This results in low cost devices for WLL as well as ensures its availability. However, such a WLL system operates in the same frequency bands standardized for the cellular or cordless applications. Therefore, allowing co-existence of both the systems, radio resource planning may be critical to avoid interference between the two systems. On the other hand, in proprietary system, frequency band is available but the licensing cost is quite high.

4.4.2 Protocol Layers and Radio Interface

The protocol layers at radio interface in WLL technology has been standardized as in other network model. However, variations are observed based on the technology used for the implementation level. Such standards/specifications define: (i) operating frequency band (ii) duplexing methodology (iii) multiple access techniques (iv) channel coding schemes

[1] It is a part of the Internet protocol suite defined by the Internet Engineering Task Force (IETF). SNMP is used in network management systems to monitor devices connected to the network for conditions that requires administrative attention.

(v) frequency modulation methods, etc. The WLL systems based on cellular and cordless mobile radio standards (GSM, DECT, etc.) follow the respective radio interface standards of GSM, DECT, etc. For example, GSM-based WLL system operates as GSM standardized by ETSI, operates in the frequency range of 890–915 and 935–960 MHz with TDMA (Time division multiple access, Section 3.2.2) and duplexing method FDD (Frequency division duplexing, Section 1.2.2). On the other hand, WLL systems based on proprietary systems are designed exclusively for WLL. Standardization of radio interface in such case is a necessity. The Air Loop proprietary system of Lucent Technologies, USA operates in 3.4–3.6 GHz frequency range with DS-CDMA (direct sequence code division multiple access, Section 3.2.3) and duplexing method FDD. The modulation technique used in Air Loop system is QPSK[2] (quadrature phase shift keying).

The protocol layers in WLL radio interface is shown in Figure 4.4. Layer 1 is the physical layer. In layer 2 functions are divided into MAC and LLC sublayers, and this layer supports acquisition and control of the radio link. The LLC supports radio interface specific features. It is implemented by link access protocol for D channel in ISDN ($LAPD_m$, Section 3.3.2).

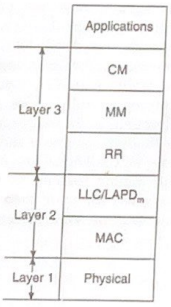


Figure 4.4 Protocol layers for radio interface.

There are three sublayers—radio resource (RR) management, mobility management (MM) and call management (CM) in Layer 3 of WLL radio interface protocol. Radio resource management is for managing logical channels such as channel assignment and performance measurement, etc. The MM, on the other hand, deals with terminal registration, location

[2] QPSK is one variant of phase shift keying described in Section 1.2.4. When BPSK uses two phases, QPSK uses four phases for coding the bits. Higher bit rates with the same bandwidth can be achieved by coding two bits into one phase shift.

update and authentication. The call/connection control, call establishment/release and management of supplementary services are done at CM sublayer.

4.4.3 Mobility Support

There are varied mobility targets in WLL-based system design. The design alternatives introduce full mobility, limited mobility or no mobility of network nodes. The full mobility enhances design cost for additional network elements for expansion in coverage area.

In full mobility, a user can roam over a wide area. It demands efficient intercell handoff mechanism while a call is in progress. The terminals in WLL system, developed around cellular technologies, are normally mounted inside a building/premises. The cellular system has sufficient link-budget[3]. The range of a WLL system based on such a cellular system is same as that of the cellular mobile system. Therefore, mobility is a feasible option for cellular technology based WLL systems.

The full mobility support requires the additional network components such as HLR (Home Location Register), VLR (Visiting Location Register), AUC (Authentication Centre) as described in Section 2.8. Further, increase in overhead is due to the management of some capabilities such as switching calls between the base stations and also between MSCs during roaming across cell boundaries. The increased cell sites and network elements supporting full mobility increase complexity in radio and network planning.

The option of limited mobility in WLL system simplifies the design of such a system. The limited mobility does not allow roaming of a node/terminal. Here the terminal is allowed to move within the coverage area of one or a few base stations (within a radius of a few kilometers). The cordless technologies, say DECT (Digital Enhanced Cordless Technology) and PHS (Personal Handy Phone System) or proprietary technologies, conventionally are not targeted for wide area mobility. Therefore, full mobility is not implementable in the WLL systems based on such cordless or proprietary technologies.

4.5 SYSTEM PLANNING

While designing a WLL system, a service provider needs to plan for radio frequency assignment, estimating number of cells partitioning the network coverage area, placement of base stations and antenna at the subscriber's end to satisfy a desired system capacity and coverage of the network.

[3] A link budget is the accounting of all of the gains (e.g. antenna gains) and losses (e.g. transmitted signal attenuation) during transmission of signal from the transmitter to the receiver in a telecommunication system. The simplest computation of received power in dB = transmitted power + gains − losses.

4.5.1 Radio Frequency Planning

The planning of radio coverage for a WLL system depends on the mobility support provided by the system. For example, a WLL system with limited or no mobility can be installed in residential areas or in office buildings. Therefore, the service provider has to plan for radio coverage targeting a small township or locality. The coverage of such an area can be ensured with a single/a number of cells forming a group centred around the township. Limited service between these groups is provided. On the other hand, the WLL system with full mobility support demands massive frequency reuse within the coverage area. Such a network is planned specially for the urban area.

Once the coverage area is decided, the area is partitioned into cells of regular shape and size. The principle of frequency reuse described in Section 2.3 is employed for planning of WLL systems with full mobility support. The relationship ($Q = D/R = \sqrt{3N}$) among the parameters frequency reuse ratio, cell radius, cluster size, etc. also holds good as in Section 2.4. The commonly chosen cluster sizes are N = 1, 3, 4, 7, 12.

The total number of radio channel provided to an operator is fixed. For a given cluster size, the number of channels for each cell is to be decided based on these two parameter values. Therefore number of channel per cell (C) = total number of channel/cluster size. For example, in a design, if the cluster size is 12 and the total number of radio channels is 480, each cell can get 480/12 = 40 channels.

The frequency planning is required for TDMA/FDMA-based system. However, for CDMA, such frequency planning is not to be done. In such a case, each cell and sector is assigned a unique Pseudorandom Noise Code (PN code) that allows a mobile station to discriminate between the signals from serving BS and the neighbouring BSs.

Increase of system capacity

In some designs, the WLL systems consider sectorized antennas at base stations. Frequency reuse is thereby increased (Section 2.10.2). Increase in frequency reuse results in increase in system capacity. Placement of sector antennas is flexible and that eases the design of WLL system. As an alternative design, cell splitting (Section 2.10.1) is used, and that increases the system capacity too.

4.5.2 Estimating Number of Cells

The network capacity and coverage define the number of cells required for the WLL system. The proposed design must satisfy the traffic capacity (A_{system}) predicted for the system. The method of estimating the number of cells to satisfy the desired system capacity is given below.

For a reasonable grade of service described in Section 2.9, and the number of channels (C) in each cell described in the previous section, the

traffic capacity per cell (A_{cell}) can be computed from the Erlang *B* table (Table 2.1).

Therefore, the estimated number of cells is (No_of_cell) = A_{system}/A_{cell}, A_{system} is the estimated network traffic capacity. However, if the No_of_cell × average cell size is less than the network coverage area, more cells are to be added.

If in a system, coverage is of prime importance, the other design methodology may be applied. In this technique, No_of_cell = Total coverage area/average cell size.

In a WLL system with limited or no mobility, roaming is not supported. Therefore, in such a case, overlapping of regions is not the necessity. On the contrary, if a WLL system supports full mobility, roaming is allowed. Overlapping of cell coverage areas is very much desirable for designing such a system. Therefore, for a WLL system supporting roaming, estimated number of cells is much higher than the computed one; whereas in WLL system without roaming facility, the estimated number of cells is equal to the computed one.

4.5.3 Deployment of Network Components

To satisfy the requirement of network coverage, a system is to be designed to provide the maximum range. That is, a system that puts priority on network coverage should place the base station antennas at high places. On the contrary, if the priority of the system is to ensure defined capacity, a smaller cell size is desirable with effective frequency reuse. Antennas are planned to be placed in lower height so that co-channel interference from other cells using the same frequencies can be minimized.

Entire design process is like a trial and change process (iterative). Initially the coverage is divided into a regular geometric shape (hexagon) cells with calculated cell size. The base stations are to be placed as close as possible to the centre of each cell. This planning can be modified considering the parameters such as availability of the installation site, local regulation (if any), cost of the location, etc.

BIBLIOGRAPHY

Clark, Martin P., *Wireless Access Networks: Fixed Wireless Access and WLL Networks—Design and Operation*, New York, John Wiley & Sons, 2000.

Huang, J., S. Tsai, Y. Lin and C. Tseng, "Design and Implementation of an OA&M System for WLL Network" *IEEE/KICS Journal Communication and Networks*, Vol. 2, No. 3, pp. 266–276, 2000.

Noerpel, A.R., and Y. Lin, "Wireless Local Loop: Architecture, Technologies and Services", *IEEE Personal Communications*, Vol. 5, No. 3, pp. 74–80, 1998.

Pandya, Raj, *Introduction to WLLs: Application and Deployment for fixed and broadband services*, New Jersey, IEEE Press & John Wiley, Inc., Publication, 2004.

Stallings, William, *Wireless Communications and Networks*, New Jersey, Pearson Education, 2006.

Stavroulakis, Peter, *Wireless Local Loops: Theory and Applications*, New York, John Wiley & Sons, Inc., 2001.

Webb, William, *Introduction to Wireless Local Loop*, 2nd ed., Broadband and Narrowband Systems, Norwood, MA, USA, Artech House, Inc., 2000.

REVIEW QUESTIONS

1. What is local loop? How is it realized in traditional telephone network?
2. How is Internet access possible through PSTN? Identify the limitations of using PSTN for Internet access.
3. What are the probable solutions to overcome the limitations of using PSTN for Internet access?
4. What are the key advantages of WLL over wired local loop?
5. What are the important parameters based on which a WLL system can be compared with a cellular mobile system?
6. Present a comparative discussion on the usage of WLL in developed countries versus that in developing countries.
7. Name two commercially-available WLL systems.
8. What are the problems that a WLL system may face arising out of spectrum usage beyond 10 GHz band?
9. What are the major components in a WLL system?
10. What are the expanded forms of IDU and ODU? In which part of the system do they belong?
11. What is the expansion of FSU? List the tasks of FSU at 'C' interface. What interface connects the PSTN with the WLL network?
12. What is the difference between 'Limited mobility' and 'Full mobility' support of WLL system? List the additional overhead needed to convert a WLL system with limited mobility to a system with full mobility.
13. (a) What are the entities to be considered during the planning of a WLL system?
    (b) During a WLL system planning, how do you decide to allocate radio channels for a cell? Explain citing an example.
    (c) What parameters would one need to estimate the number of cells required for a WLL system? Describe the method to estimate the number of cells for a WLL system based on those parameters.

5

WIRELESS LOCAL AREA NETWORK

This chapter explains wireless LAN (WLAN) including its standards and applications. A WLAN connects network components through a wireless medium over a relatively short distance. The wireless medium is typically radio or infrared. Unlike WAN (wide area network), the WLANs are normally owned, controlled and managed by a single person or organization. The span of a WLAN usually can be a building or a place even smaller than that. In practice, WLAN coexists with LAN within an organization.

5.1 APPLICATIONS

The WLANs are effective in an environment where the cable installations are either not cost-effective or not at all practical, e.g. at a make-shift fare ground, or at a heritage buildings with insufficient twisted pair and where drilling holes for new wiring is prohibited. WLAN can be used as an alternative to wired LAN. However, in most organizations, WLAN may co-exist with LAN and they serve the requirements of the organization in a collaborative manner. For example, a factory can have an office area that are to be separated from the factory floor and must be linked to it for networking. In this environment, the office may have a wired LAN that is connected with WLAN installed in the factory floor.

5.2 DATA TRANSFER

WLAN supports both the asynchronous and synchronous data transfer. The asynchronous data transfer suits the applications that are not time

sensitive (e-mail, file transfer, etc.); whereas the synchronous mode supports time-bound applications like video and packetised voice. A WLAN's transmission range is affected by the characteristics of the frequency band it uses. On an indoor IEEE 802.11b WLAN, it is usually possible to maintain a speed of 11 Mbps at distances up to 100 ft.

5.3 WLAN CATEGORIES

WLAN is categorized according to its transmission techniques, or according to its connectivity of nodes.

5.3.1 Transmission Technique-based Categorization

According to the transmission technique used, there can be three types of WLAN.

- *Infrared LAN (IRLAN):* It employs infrared for transmission. An individual cell of an IRLAN is limited to a single room, because infrared light does not penetrate opaque walls.
- *Spread spectrum LAN:* This class of LAN makes use of spread spectrum transmission technology. In most cases, these LANs operate in the ISM (Industrial, Scientific and Medical) bands to avoid FCC (Federal Communications Commission) licensing wherever applicable.
- *Narrowband microwave:* These operate in microwave frequencies but do not use spread spectrum. Some of these products operate at frequencies that require FCC licensing.

5.3.2 Connectivity-based Categorization

The infrastructure WLAN and Ad Hoc WLAN are differentiated according to the connectivity of nodes.

Infrastructure WLAN: Such a WLAN must have at least one wireless computer (station/client) and one wireless control module (base station/access point). The Access Point (AP) is connected at one end with the wired network and at the other end, it is wirelessly connected with the client. The client uses the AP to access the resources of a traditional wired network.

Ad Hoc WLAN: Wireless clients are connected in peer-to-peer mode. They communicate directly with each other without the use of a wireless AP. It is basically a short-lived network, created for a particular purpose. WLAN installed within a conference venue is an example of Ad Hoc WLAN.

5.4 THE WLAN STANDARDS

The standards that primarily have emerged are the IEEE 802.11 in the US and High Performance European Radio LAN (HIPER LAN) in Europe.

The IEEE standard 802.11 specifies the physical and MAC layers for operation of WLANs and employs the direct sequence spread spectrum (DSSS) and frequency hopping spread spectrum (FHSS) technologies. The standard applies to 2.4 GHz band. It is mainly devised to support asynchronous data transfer mode and optionally has synchronous data transfer mode as well. The following subsections describe the basics of WLAN supported by the IEEE 802.11.

5.4.1 Architecture

The 802.11 supports both the infrastructure and Ad Hoc WLAN. In 802.11 infrastructure WLAN, the area covered by a network is divided into cells. Each cell is a Basic Service Set, controlled by the base station/access point. The service set identifier (SSID) differentiates a WLAN from the other. It is a 32-character unique identifier attached to the header of packets sent over a WLAN. The SSID acts as the password when a wireless client tries to connect to the Basic Service Set. All APs and all the clients attempting to connect to a specific WLAN must use the same SSID. That is, an SSID corresponds to a network name. It allows logical separation of WLANs.

A device (AP, client device) will not be permitted to join the Basic Service Set unless it can provide the unique SSID. Even though a single

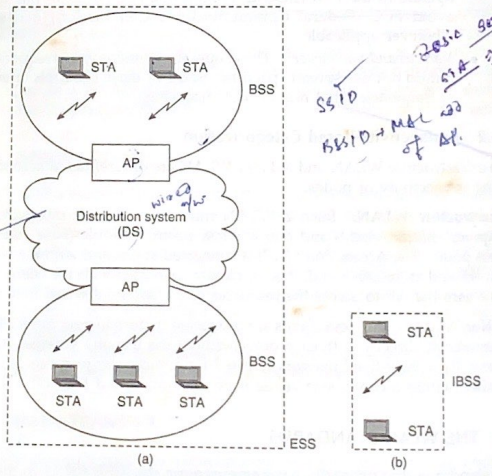


Figure 5.1 The IEEE 802.11 WLAN architecture.

cell, with a single AP, can form an infrastructure WLAN, most installations consider several cells. The APs are connected through backbone called Distribution System (e.g. Ethernet). An entire WLAN, including its cells, respective APs and the Distribution System (described later) is viewed by the upper layers of OSI model as a single 802 network and is referred to as Extended Service Set. The basic architecture of IEEE 802.11 WLAN is shown in Figure 5.1. The brief descriptions on the functionality of its major components are noted below.

Station (STA): A wireless STA (Stations) contains an adapter PC card[1] or an embedded device to provide connectivity to the wireless medium. It consists of protocol stack having MAC (Medium Access Control) and PHY (Physical) layers responsible for communication. An STA can be mobile or stationary and can support a number of services such as authentication/deauthentication, data delivery and privacy. Before getting access to WLAN, each STA should authenticate itself without which the use of WLAN is not allowed for data delivery. On the other hand, once an STA is deauthenticated, it can no longer access the WLAN. However, the STA can provide data protection in wireless medium, only to a limited extent.

Access Point (AP): It functions as a bridge between the STAs and the existing network backbone. In an infrastructure BSS, all mobile STAs communicate with the AP. The AP provides both the connection to the wired LAN (if any) and the local relay function for the BSS. Thus, if one mobile STA in the BSS communicates with another mobile STA (the target STA), the communication is sent first to the AP and then from the AP to the target mobile STA.

Basic Service Set (BSS): A BSS is a set of STAs that communicates among themselves. There are two classes of BSSs—one includes AP, and is called infrastructure BSS, and the other class that contains only the mobile STAs, called independent BSS (IBSS).

Infrastructure BSS: The infrastructure BSS consists of a single AP and one or multiple wireless clients (STAs). All the STAs in a BSS communicate through the AP. The AP provides connectivity to the wired LAN. This is known as Infrastructure WLAN.

Independent BSS (IBSS): In IBSS, there is no AP, so the mobile stations do not have a connection to wired network. As there is no AP, the mobile STAs communicate directly among themselves. This may restrict a mobile STA to communicate with all other mobile STAs even though these are part of the same IBSS. Further, there is no relay function in an IBSS. Thus if two STAs are to communicate with each other, these two must be in direct communication range. This is known as Ad Hoc WLAN.

[1] An interface standardized by the Personal Computer Memory Card International Association.

Extended Service Set (ESS): An ESS is a set of two or more APs connected to the same wired network. The same network means a single logical network segment bounded by a router.

Distribution System (DS): The APs of BSSs are interconnected via the DS (distribution system). This allows mobility (roaming) of STAs from one BSS to another. The DS is a logical component. The specification does not put any restriction on how DS is implemented. It only specifies the services put, namely association, reassociation, disassociation, distribution and integration.

Association: The association service is invoked to make a logical connection between a mobile STA and an AP. The connection is necessary for the AP to accept data frames from the mobile STA and then to allocate resources to the mobile STA.

Reassociation: Whenever a mobile STA identifies disconnection from the servicing AP, a reassociation service is invoked by the STA to associate itself with a new AP. Once the mobile STA is reassociated with a new AP, it contacts the formerly associated AP to obtain frames that are waiting (if any) there for delivery to the mobile STA.

Disassociation: This service is used either by an AP or by a mobile STA. When demand exceeds with respect to available resources, an AP uses this service and informs a number of users that it can no longer provide the logical connection to the WLAN and forces them to reassociate to a different AP.

When a mobile STA no longer requires its connection to the WLAN, it invokes disassociation service. This requirement arises due to switching off the STA, disconnection of STA as WLAN node, etc. Upon receipt of the disconnection request, an AP frees the resources provided to the STA.

Distribution: This service is also invoked by a mobile STA or by an AP. When an STA belonging to a BSS desires to exchange information with an STA belonging to another BSS, the information is sent to the AP of the source BSS that transfers it to the DS. The DS in turn transfers the information to the AP of destination BSS and the AP sends the same to the destination mobile STA.

Integration: With the help of this service an STA in a WLAN can communicate with a station in wired LAN connected with DS. All the necessary conversions required to exchange between wireless and wired LAN are taken care of by the service.

WLAN configuration

The parameters for configuring WLAN are SSID, channel number and security setting. The task of configuring a client/STA is performed with the help of configuration utilities/tools provided by the manufacturers. All the clients and the AP (present in an infrastructure WLAN only) must

use the same parameter settings for a WLAN to function correctly. A client must be configured either in Infrastructure or in Ad Hoc mode to join an infrastructure or an Ad Hoc WLAN respectively. The parameter SSID setting is important to identify a WLAN/BSS to which the client wants to join. SSID is broadcast either by AP (infrastructure WLAN) or by other clients (Ad Hoc WLAN) and the client's adapter card identifies it automatically. For security setting, a client must be configured with WEP (wired equivalent privacy) key, an encryption key. The encryption key of the client must match that of the AP.

IEEE 802.11a and 802.11b standards transmit signal in a narrow radio frequency range of 2.4 GHz. This signal range is divided into a number of channels. During configuration, to avoid interference from other electronic devices such as cordless phone, microwave oven, etc. in the vicinity, setting of channels is essential. In case of Infrastructure WLAN, APs have multiple channels on which they broadcast signals. If a WLAN encounters any interference from other nearby WLANs or any other devices, channel setting can be changed.

5.4.2 Protocol Stack

The WLAN architecture follows OSI network model. In OSI, protocol layers (Network to Application) are independent of network architecture. Therefore, the 802.11 standard provides specifications to address the lower layers of the OSI model, that is, specification to the physical (PHY) layer, medium access control (MAC) and logical link control (LLC) layer. The said lower layers of the protocol stack are shown in Figure 5.2. The functions of these three layers are:

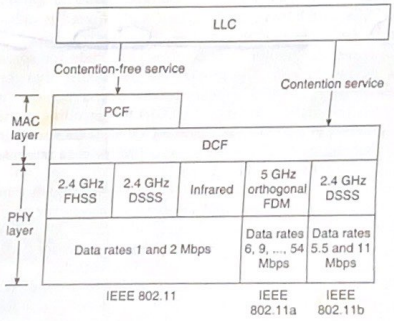


Figure 5.2 Protocol stack.

- PHY layer
  - converts signals from the MAC layer to radio signals and vice versa.
- MAC layer
  - assembles (during transmission) data received from the upper layers into a frame with address and error detection fields.
  - disassembles (during reception) data frame received from the PHY layer and performs address recognition as well as the error detection.
  - monitors access to the LAN transmission medium.
  - ensures error free transmission.
- LLC layer
  - provides interface to upper layers.

PHY layer: The Physical layer gives the specification of transmitting and receiving raw bit stream. It converts communication requests from the upper layers into hardware-specific operations for transmission of radio signals. To convert the signal, it performs signal encoding/decoding and modulation/demodulation during transmission and reception respectively. A PHY layer has two sublayers. One deals with the wireless transmission medium such as signal encoding, decoding, transmission, reception, etc. The other sublayer sets the interface with the MAC including preamble creation/elimination for synchronization.

The IEEE 802.11 Physical (PHY) layer communicates in three different modes—(i) FHSS (Frequency Hopping Spread Spectrum) in 2.4 GHz band, (ii) DSSS (Direct Sequence Spread Spectrum) in 2.4 GHz band and (iii) IR (Infrared–frequency range is close to terahertz). FHSS and DSSS modes WLAN uses radio signal and IR mode WLAN uses Infrared signals for communication. The primary difference between radio signal and Infrared is radio operates in GHz band whereas IR operates in close to terahertz range.

As discussed in Chapter 3, DSSS uses chip code for spreading the signal. The longer the chip, the greater the probability that the original data can be recovered. DSSS appears as low-power noise to a receiver other than the destination receiver. In DSSS for encoding signal DBPSK (Differential Binary Phase Shift Keying) and DQPSK (Differential Quadrature Phase Shift Keying) is used for 1 Mbps and 2 Mbps data rate transmission respectively. DSSS uses 5 MHz channel.

FHSS uses a narrow band carrier that changes frequency in a pattern known to both the transmitter and receiver. To a receiver other than the destination one, FHSS appears as impulse[2] noise. For 1 Mbps FHSS system two level GFSK (Gaussian Frequency Shift Keying) and for 2 Mbps system four level GFSK are used. FHSS uses 1 MHz channels.

Both DSSS and FHSS operate in 2.4 GHz ISM band. For both the DSSS and FHSS, the availability of channels depends on the allocated bandwidth by frequency regulatory agencies of respective countries. For

[2] An impulse noise is a short burst type noise that usually lasts for less than one second.

example in Europe 13 number of channels, in the US 11 number of channels and in Japan one channel is provided for DSSS. In FHSS, the number of hops is 79 in the US and Europe and number of hops available ranges from 23 in Japan. FHSS systems stay at each hop for 20 msecs. However, only a few channels can be effectively deployed in close proximity to one another. Figure 5.2 shows different data rates for different standards such as 1 and 2 Mbps data rates are for IEEE 802.11.

IR cannot penetrate opaque objects and therefore an IR based WLAN operates either in directed (line-of-sight) or diffuse/reflective mode. IEEE 802.11 defines 802.11 IR that operates by reflecting/bouncing light from ceiling and walls to provide connectivity within a room or small office. It provides data rate 1 to 2 Mbps. As IR cannot penetrate walls, eavesdropping on the system from outside of the system is not possible and therefore it is highly secured. Moreover, there is no interference from other systems that operate in radio frequency. In spite of providing secured, interference-free operation there are no products available that are 802.11IR compliant. The reason is probably the lack of conformance with standards and lack of regular update that is essential to compete with the standards using radio signal for operation.

The PHY layer creates/eliminates some bits as preamble to add with data to synchronize transmitter and receiver for transmission and reception respectively. There are two types of preambles—one is long and the other is short preamble to support different types of services. The long preamble is mandatory for a device to provide data services. The optional short preamble is provided in the standard to improve the efficiency of throughput of network during transmission of real time sensitive services such as voice, voIP (voice over IP) and stream video.

The IEEE 802.11b provides up to 11 Mbps raw data rate with Quadrature Phase Shift Keying (QPSK) modulation. In addition, it defines a dynamic rate shifting scheme that allows data rates to be automatically adjusted considering the noise margin. In other words, a device transmits at lower speed when the noise level is high. The device automatically shifts to a high speed mode if the noise condition improves.

MAC layer: This layer of the protocol stack provides interface between the physical layer (PHY) and the upper layers. The major tasks done by a MAC are the error correction, transmission of data and control of access mechanism to the shared radio medium.

Error-free transmission: The IEEE 802.11b MAC defines two features for error control: cyclic redundancy check (CRC[3]) and packet fragmentation. CRC ensures that the data is not corrupted in transit. Whereas, packet

[3] CRC is a powerful error-detecting code in communication system. One implementation is the transmitter generates n-k bits to transmit a data of k bit such that the data bit is converted to n bits divisible by a predetermined number known by the receiver. The receiver divides received data by the number; if the results show a remainder, error is detected in transmission.

fragmentation allows a large packet to be broken up into small pieces while transmitting. This reduces the chances for retransmission as the probability of a packet getting corrupted decreases with smaller packet size. The small packets also reduce the retransmission time when the data are corrupted as the smaller packets can be retransmitted more quickly.

Access mechanism: Access control is the ability to permit or deny the use of physical transmission medium on a communication network. Similarly, WLAN being a communication network, also has access control mechanism so that clients get access of shared radio medium without any conflict. WLAN supports both the contention-based and contention-free mechanisms. In contention-based mechanism, clients compete to gain control of the transmission medium for transmission of each packet. Whereas in contention-free mechanism, there is no such competition. Each client gets a time slot to transmit. The physical transmission time is divided into cycles and each cycle is again divided into two slots, a contention period (CP) and a contention-free period (CFP). This arrangement ensures medium access to both asynchronous as well as synchronous traffic in each cycle. The contention-based strategy is efficient for lightly-loaded system. As the load increases, competing clients prevent each other from gaining control of the medium and hence data transmission is impaired and packet delay increases. The contention-free strategy is useful for comparatively loaded system, e.g. transmitting voice, video.

The IEEE standard, as shown in Figure 5.2, specifies one mandatory contention-based medium access control method and two optional ones. The mandatory one is Distribution Coordination Function (DCF). This is implemented by CSMA/CA (carrier sense multiple access/collision avoidance). To enhance the performance of CSMA/CA, one optional method Ready to Send (RTS)/Clear to Send (CTS) mechanism is defined. The other optional method is Point Coordination Function (PCF), a contention-less medium access control.

In DCF, the decision to transmit is distributed over all the nodes using carrier sense mechanism. On the other hand, the PCF follows centralized access, which involves regulation of transmission by a centralized decision maker. The AP plays the role of centralized decision maker or point coordinator. As there is no central decision maker like AP in Ad Hoc WLAN, only the distributed access control is used for ordinary, asynchronous, bursty type of traffic such as file transfer. On the contrary, both the centralized and distributed accesses are applicable for Infrastructure WLAN attached to a backbone wired LAN. The centralized access is especially useful if some of the data is time sensitive or high priority such as voice, video. In a word, the DCF supports asynchronous data transfer and optional synchronous data transfer is made possible by PCF. The DCF and PCF are described in the following subsections.

Distribution Coordination Function (DCF): The CSMA used in DCF is a well-known protocol used in Ethernet. In CSMA, a station desiring to transmit senses the medium. If the medium is busy indicating that some other station is transmitting, the station will defer its transmission. If the medium is free, the station is allowed to transmit. In this scenario, there is a probability that multiple stations may sense the medium is free and start transmitting data simultaneously resulting a collision. In the Ethernet protocol, used in wired LAN, the collision is detected by the transmitting stations and then retransmits the same. However, in WLAN the collision detection can't work due to the following problems:

- *Hidden station problem:* Let us consider the scenario of Figure 5.3 and assume that A is transmitting data to B. Further, node C desires to transmit data to B and tries to sense the medium. C, being out of the range of A, cannot sense the medium. Therefore, C starts transmitting to B and creates interference at B.
- *Exposed station problem:* Let us consider B is transmitting to A and simultaneously C desires to transmit data to D. C then tries to sense the medium and concludes that the medium is busy and postpones the intended transmission to D. However, the transmission could be possible as data transmission from C to D does not result in collision with the ongoing transmission between A and B.

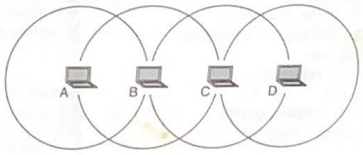


Figure 5.3 *Hidden/exposed station problem.*

The above two limitations of WLAN can be arrested by providing full-duplex radio capable of transmitting and receiving at once with a high price.

The WLAN standard utilizes a collision avoidance mechanism. As soon as a node receives a packet to be sent, it checks whether any other node is transmitting at the same time. If the channel is clear, then the packet is sent. If the channel is not clear, the node waits for a randomly chosen period of time, and then checks again to see if the channel is clear. Thus the basic principles of CSMA/CA are listen before talk and contention. This is an asynchronous, connection-less[4] message passing mechanism,

[4] In connection-less mode of service, the data is transmitted to the other without ensuring that the recipient is available and ready to receive the data. It does not require a session connection between sender and receiver. The other mode of service is connection-oriented that requires a session connection be established before any data can be sent.

delivering the best effort service. CSMA/CA is derived from CSMA/CD which is the base of Ethernet. On a wire, the transceiver has the ability to listen while transmitting and so to detect collisions (with a wire all transmissions have approximately the same strength). But, even if a radio node could listen on the channel while transmitting, the strength of its own transmissions would mask all other signals on the air.

CSMA/CA scheme

Following are the algorithmic steps that are to be followed in CSMA/CA scheme.

- A station willing to transmit senses the medium
- if the medium is idle
  - wait for a predetermined time (delay IFS)
    /* IFS–Interframe space */
  - if the medium is still idle
    - the station transmits
  - else
    - wait until current transmission ends ——L1
    - wait for a predetermined time (delay IFS)
    - if the medium is still idle
      - backs off a random amount of time —L2
      - if the medium is idle
      - the station transmits
      - else
        - repeat from L2
    - else
      - repeat from L1
- else repeat from L1

Back off is a well-known method to resolve competition among different stations willing to access the medium. The method requires each station to choose a random number between 0 and a given number, and waits for this number of slots before accessing the medium so that no other station accesses the medium simultaneously. Repeated unsuccessful attempts to transmit data result in longer back off times. Without such a back off the following situation can occur—two or more stations attempt to transmit at the same time, causing a collision. These stations then immediately attempt to retransmit, causing a new collision.

MAC retransmissions

WLAN using the IEEE 802.11 physical and MAC layers is subject to considerable unreliability. There is a higher error rate on the air than on a wire. So the chances of corrupted packets in WLAN are more in comparison to the data transmission in wired medium. TCP is not designed to take care of packet losses at the MAC layer. Moreover, the problem with the CSMA/CA protocol is that the transmitter cannot detect collisions on the

medium. That is why, most of the MAC protocols implement positive acknowledgement and MAC level retransmissions to avoid loss of packets. However, as the broadcast and multicast packets are not acknowledged, the scheme may not be workable for such cases. The IEEE 802.11 implements the similar scheme augmented with a frame exchange protocol.

RTS/CTS mechanism

The fundamental problem with carrier sense is that the transmitter tries to estimate whether the channel is free at the receiver considering only the local information. To enhance the reliability of CSMA/CA further, a four-frame exchange, namely Request to Send (RTS)/Clear to Send (CTS) mechanism is used. The name of the changed protocol is MACA (Multiple Access with Collision Avoidance) protocol. The RTS/CTS is also called virtual carrier sense. It is a handshaking[5] mechanism. Prior to sending a packet, the transmitter sends a RTS and waits for a CTS from the receiver. The reception of CTS indicates that the receiver is able to receive the RTS. The RTS then alerts all stations that are within the reception range of the destination, informing that an exchange is underway. The RTS and CTS messages contain the size of the expected transmission, i.e. how long the transmission can last. All the nodes avoid accessing the channel after hearing the CTS even if their carrier sense indicates that the medium is free. This ensures the collision avoidance in a WLAN. The MACA can avoid the problems of hidden and exposed station described earlier.

MACA

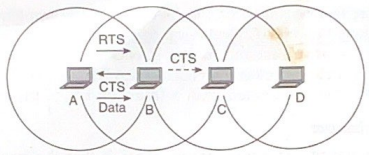
Hidden Station Problem avoidance (Figure 5.4 (a))

- A sends a RTS frame to B along with source and destination names and length of future transmission (A, B, L); C does not hear RTS.
- B acknowledges by sending a CTS frame to A containing (A, B, L) where length of future transmission (L) is copied from RTS; C hears CTS and must not send anything to B for L time duration.
- A sends data.

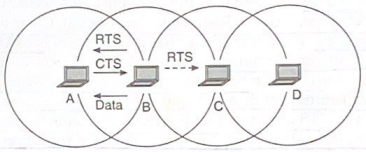
Exposed Station Problem avoidance (Figure 5.4 (b))

- B sends a RTS frame to A along with (B, A, L); C listens this RTS indicating medium busy. This indication does not matter for C to start transmission to D.
- A acknowledges by sending a CTS frame to B along with (A, B, L); C does not listen this CTS.
- B sends data.

[5] Handshaking is an exchange of information over a communication network that establishes a valid connection between the two stations. A common example of handshaking is: a modem dials up to a computer network and negotiates the parameters, such as baud rate and error correction, to establish connection.



(a) Hidden station problem



(b) Exposed station problem

Figure 5.4

Point Coordination Function (PCF): As described earlier, the point coordination function is applicable for sending time-sensitive data, e.g. voice, video in infrastructure WLAN with wired backbone. The base station/AP transmits a beacon to announce the CFP to all clients. The clients in turn go to hold state and stop DCF mode transmission. The base station/access point polls[6] the clients following a predetermined strategy. In polling the AP sends a poll packet to trigger the transmission by the client. Once a client is polled, it has the right to transmit while all other clients remain idle. Receiving a poll packet, a client can request a connection or transmit packet depending on the connection-oriented or connectionless implementation of polling. The base station retains total control over the transmission medium allowing transmission of packets having variable size.

The point coordinator can seize the medium and lock out all asynchronous traffic while it issues polls and receives responses. In the worst case, it can lock all asynchronous traffic by repeatedly issuing polls. To get rid of this, as described earlier, it introduces the concept of contention period and contention-free period. In contention-free period, it issues polls and for the contention period it allows asynchronous access.

Summarily the IEEE 802.11 offers a polling channel access mechanism, i.e. PCF in addition to the DCF implemented by CSMA/CA. It provides a distributed access control mechanism with an optional centralized control built on the top of that. The lower sublayer of MAC is DCF and upper

[6] Polling is the process in which an issuing node (polling station) of a communication network broadcasts a query to every other node in the network and waits to receive a unique response from each of them.

sublayer is PCF. The DCF uses a contention algorithm to provide access to all the traffic. Ordinary asynchronous traffic directly uses DCF. The PCF, a centralized MAC algorithm, provides contention-free services such as voice, and video.

5.4.3 Roaming in WLAN

Roaming means a wireless client which can be able to switch between network access points (APs) seamlessly. The roaming occurs due to physical movement of a client node from one cell (BSS) to another or due to load balancing between the access points. However, in both the cases, the roaming process allows a client to be attached to a new AP on an appropriate channel. This calls for AP to AP communication. The important part of the roaming process is to inform rest of the network that the client is shifted to a new AP. This was first implemented in 802.11f, by the Inter-Access Point Protocol (IAPP), as a trial-use standard. The client's identity and BSS to which it is currently associated are conveyed to other APs by the IAPP. The new AP broadcasts the client's association to it so that the DS updates its MAC address table database to record the client's new location. The DS informs the old AP about the client's disassociation.

If a client under the control of access point AP-1 physically gets closer to a new access point AP-2, the signal strength received by the client from AP-1 drops, whereas the signal strength from AP-2 received by the client rises. When the signal strength from AP-2 is stronger than that from AP-1, the client is made to roam to AP-2. Figure 5.5 depicts the roaming of client C from AP-1 to AP-2.

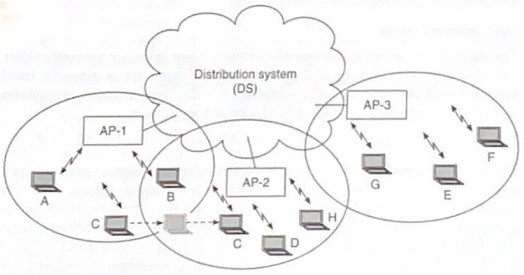


Figure 5.5 Roaming.

While roaming takes place due to load balancing of APs, the APs negotiate evaluating the parameters such as the number of associated clients on different APs, the signal strength of the client received by APs,

the traffic loads at different APs, etc. Based on the information, ESS finds out which APs are heavily loaded. Accordingly the decision for intra-ESS load balancing is taken. Load balancing is implemented either by not allowing a new client/STA to access a heavily-loaded AP or by applying forced roaming on a selected number of STA.

To ensure that a client can be able to roam seamlessly, it is necessary for the APs to maintain –(i) connection to the same IP subnet so that the client won't have to change the IP address, (ii) the same SSID to identify the wireless network and (iii) the same authentication and encryption schemes that are already known to the client. The 802.11 standard does not define how roaming should be performed but it defines various services such as association, re-association etc. described in Section 5.4.1. The association/re-association processes are implemented as noted below.

All APs periodically send beacon frames. In between beacon frames, the stations are in power-saving mode. If a station joins the network or decides to change its AP, it sends a probe frame. All APs within the range of the station reply with a probe response frame. The station selects an AP and sends an association request frame. The selected AP then replies with an association response frame.

5.4.4 WLAN Security

Security in terms of user authentication and data privacy is provided in a WLAN. As the communication in WLAN is broadcast in nature, it is essential to authenticate the user/station to prevent unauthorized access to the network resources. It must also prevent data tampering for maintaining integrity and privacy of transmitted data.

User authentication

The user/client authentication means wireless client or device authentication. The Service Set Identifier (SSID) and MAC address are commonly used for such authentication. Further, the IEEE 802.11 specification stipulates two mechanisms—open and shared key authentication.

SSID authentication

The SSID is a configurable identification that allows a client to communicate with an appropriate access point. With proper configuration, only the client with correct SSID can communicate with the APs. The SSID acts as a shared password between the APs and clients. The SSID is not designed, nor intended for use, as a security mechanism. It is advertised in plain text in the AP beacon messages. Although beacon messages are transparent to the users, an eavesdropper can easily determine the SSID with an 802.11 WLAN packet analyzer. To put it differently, to gain access to a WLAN, a client must be configured with the appropriate SSID. However, the SSID neither provides any data privacy function nor does it truly authenticate the client to the access point.

MAC address authentication

MAC address authentication verifies the client's MAC address against a locally configured list of allowed addresses. MAC authentication is used to augment the open as well as the shared key authentication provided by the IEEE 802.11 and therefore, reduces the probability of access to a network by an unauthorized device.

Open authentication

Open system authentication provides a mechanism to exchange identities and agreement while exchanging data between the two stations (STAs). A station STA-1 sends a MAC control frame, known as authentication frame, to another station STA-2. The station STA-2 (if agrees) responds to STA-1 with its own authentication frame. The access point grants any such request for authentication.

Shared key authentication

In shared key authentication, the client configures a static wired equivalent privacy (WEP) key. The WEP keys can function as a type of access control. A client without the correct WEP key cannot send or receive data to and from an AP. The WEP encryption scheme adopted by the IEEE 802.11 provides encryption with 40 bits or 104 bits of key. It is based on the RC4[7] algorithm. The encryption keys must match on both the client and the AP for frame exchanges to succeed. The steps of authentication are:

- The client sends a MAC authentication frame including the client identification and shared key as an authentication request to the AP, requesting shared key authentication.
- The AP responds with an authentication response message containing the challenge[8] text. This challenge text is generated using WEP PRNG (pseudorandom number generator).
- The client uses its locally configured WEP key to encrypt the challenge text and reply with a subsequent authentication request.
- If the AP can decrypt the authentication request using WEP and the secret key, shared with the client, and if that results in retrieval of the original challenge text, then it responds granting the client access.

[7] RC4 algorithm is a shared key stream cipher algorithm requiring a secure exchange of a shared key. A stream cipher is a method of encrypting text in which a cryptographic key and an algorithm are applied to each binary digit in a data stream, one bit at a time. The algorithm is used identically for encryption and decryption as the data stream is simply XORed with the generated key sequence.

[8] It is a part of challenge-based authentication procedure that takes place between two nodes of a computer network to be authenticated. One node seeks a challenge/question and the other node replies with response/answer.

In shared key authentication, the client uses a pre-shared WEP key to encrypt challenge text sent by the AP. The AP authenticates the client by decrypting the shared key response and validating the challenge text. The process of exchanging the challenge text occurs over the wireless link and hence an eavesdropper can capture both the plain text (challenge text) and the cipher-text response. The following are execution steps required for a client to get access to an AP:

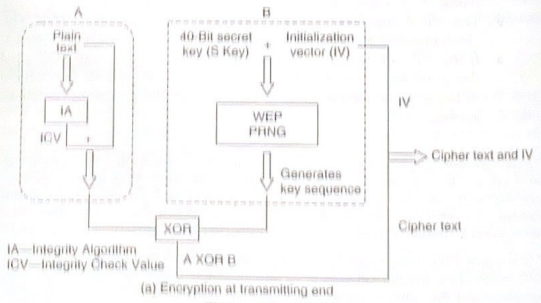
- Client broadcasts a probe request frame on every channel.
- APs within the range respond with a probe response frame.
- The client decides which AP is the best for access .
- It sends an authentication request to the selected AP (this authentication process is either done by an open or by shared key authentication).
- The AP sends an authentication reply.
- Upon successful authentication, the client sends an association request frame to the AP.
- The AP replies with an association response.

The client is able to pass traffic to the AP.

Data privacy

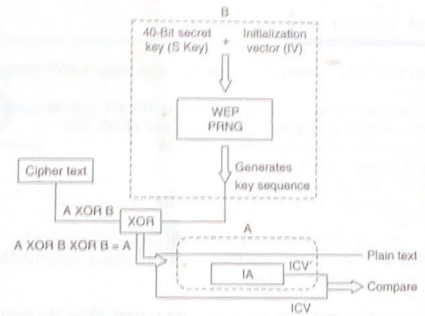
IEEE 802.11 provides WEP-based moderate level of security. It uses an encryption algorithm based on RC4 encryption scheme. Figure 5.6 describes the encryption process.

A 40-bit secret key together with an initialization vector (IV) is fed as input to a WEP PRNG (pseudorandom number generator) defined in RC4. The method exploits the property P XOR Q XOR Q = P. It means plain text is XORed with the key sequence in the transmitting side. In the receiving side the output of the transmitting end is again XORed with the key sequence to get back the plain text.



(a) Encryption at transmitting end

Figure 5.6 *Contd.*



(b) Decryption at receiving end

Figure 5.6 WEP.

BIBLIOGRAPHY

Bantz, D., and F. Bauchot, "Wireless LAN Design Alternatives", *IEEE Network*, 1994.

Crow, B., et al., "IEEE 802.11 Wireless Local Area Networks", *IEEE Communications Magazine*, Vol. 35, No. 10, 1997.

Geier, J., *Wireless LANs*, New York, Macmillan Technical Publishing, 1999.

Kleinrock, L., and F.A. Tobagi, "Packet switching in radio channels: part I–carrier sense multiple access modes and their throughput-delay characteristics", *IEEE Transactions on Communications*, Vol. 23, No. 12, pp. 1400–1416, 1975.

Lam, S.S., "A Carrier Sense Multiple Access Protocol for Local Networks", *Computer Networks*, Vol. 4, pp. 21–32, 1980.

LaMaire, R.O., et al., "Wireless LANs and Mobile Networking: Standards and Future Directions", *Communications Magazine*, Vol. 34, No. 8, August 1996.

Ohara, B., and A. Petrick, *IEEE 802.11 Handbook: A Designer's Companion*, New York, IEEE Press, 1999.

Pahlavan, K., T. Probert and M. Chase, "Trends in Local Wireless Networks", *IEEE Communications Magazine*, 1995.

Santamaria, A., and F. Lopez-Hernandez, *Wireless LAN systems*, Boston, Artech House, 2000.

Stallings, William, *Wireless Communications and Networks*, New Jersey, Pearson Education, 2006.

REVIEW QUESTIONS

1. Can WLAN be considered an alternative to wired LAN? Justify your answer.
2. Write a few areas of application where WLAN is most desirable.
3. What are the data transfer characteristics of WLAN?
4. Describe the following types of WLAN:
   (a) Infrared LAN
   (b) Spread spectrum LAN
5. (a) Explain the role of access point (AP) in a WLAN.
   (b) What is STA?
   (c) What is BSS? How do you classify it?
6. What are the services specified for Distribution System? Describe briefly each of the services.
7. How does Ad Hoc WLAN work?
8. What are the parameters for configuring a client? Write the significance of each parameter.
9. (a) What are the alternative communication specifications at physical layer?
   (b) Write the tasks of MAC layer.
   (c) What is CRC?
10. (a) What is meant by access control?
    (b) What are contention-based and contention-free access control mechanisms?
    (c) Why does CSMA/CD not work for medium access control of WLAN?
    (d) What is the alternative mechanism to CSMA/CD? Write steps of operation to implement such an alternative mechanism.
11. Give a brief specification of WLAN supported by IEEE 802.11.
12. (a) Describe briefly the hidden station and the exposed station problems.
    (b) What is MACA?
    (c) How are the problems of hidden station and exposed station solved using MACA?
13. For what types of applications PCF is used and for what types of applications DCF is used?
14. What are the possible causes of roaming? How roaming occurs due to load balancing?
15. What characteristics of a wireless LAN present unique security challenges which are not found in wired LANs?
16. (a) What level of security can WLANs provide?
    (b) Draw the block diagram showing encryption/decryption process applied for data privacy in WLAN. Describe the steps of operation.
    (c) How are SSID and MAC addresses used to authenticate a client?
    (d) What is shared key authentication? Write steps of operation in performing such authentication.
    (e) How does a client get access of an AP?

6

WIRELESS DATA SERVICE

In parallel with the service of voice over wireless network, demand for data service on move has also been growing at a very rapid rate. The data service on move is meant for the services such as file transfer, Internet access, etc. from within wireless network. The two players behind the fulfilment of this demand are the easy availability of portable devices such as Laptop, PDA, Cell phone and the services available for providing Internet access. Getting Internet access from any location at any time with portable devices is today's minimum requirement. Sometimes even the requirement increases further for having uninterrupted Internet service while the device changes its location during a session.

As mentioned in Chapter 1 (section 1.3), wireless networks are categorized as wireless LAN, MAN and WAN. The major focus of this chapter is to pay attention on the Internet access and other data services on move from within wireless WAN. The motivation behind considering wireless WAN is to provide mobile data service with wide coverage.

Conventional cellular radio and landline telephony use circuit switching. It dominates the public switched telephone network or PSTN. On the contrary, packet switching dominates data networks like Internet. In packet switching, data is split into several packets that are routed towards destination. The different routes followed by the packets, targeted to a destination, may reach the destination at different times. At the destination, the packet assembler arranges those packets in order. This approach can be acceptable when surfing a web page or downloading a file, since a marginal delay is hardly noticed in such applications. However, even a minor delay in image, video or voice transfer is undesirable and the delay causes service quality to fall. The circuit-switching mechanism guarantees

the quality of such time-sensitive services as all packets in circuit switch go in order.

The main benefit of packet-switched radio access is that it reserves radio resources only when there is something to send. The same radio resource is shared by all mobile stations within a cell, and thereby ensures effective use of scarce resources. Moreover, in contrast to time-oriented charging applied for circuit-switched connections, packet-switched data services allow charging depending on the amount of data transmitted and the quality of service negotiated. For example in circuit-switched network, for web browsing, one has to pay for the entire duration, although the channels remain idle for most of the time.

6.1 FIRST INITIATIVE ON DATA SERVICE

From the point of view of mobile voice communications, 2G cellular networks have provided quite sufficient and satisfactory services to users. Moreover, the 2G networks have successfully provided the low data rate services in the form of the popular short message service (SMS). In the post-2G scenarios, data services at very high speed and Internet access on move are considered as the main service components. In other words, the main goal of the post-2G systems is to provide a reliable wireless data and Internet service to mobile users. Since the time getting data services over wireless became essential, attempts have been made to provide such services based on the existing cellular network. HSCSD (High Speed Circuit-Switched Data) and CDPD (Cellular Digital Packet Data) are considered as the first initiatives for providing wireless mobile Internet services over cellular networks.

Although the circuit switch is not suitable for data services, GSM being a circuit-switched network, in its post-2G implementations, provides the data service called HSCSD. The GSM, having 9.6 kbps data rate, provides a higher rate of data service through HSCSD by combining several channels. On the other hand, the CDPD is a specification for supporting wireless access to the Internet and other public packet-switched networks over cellular telephone networks. It supports TCP/IP and Connectionless Network Protocol (CLNP). CDPD is the best-known packet-switched service designed for operation with an analog cellular system, specifically AMPS (Advanced Mobile Phone System). Although CDPD can be implemented as a stand-alone service, it is almost always implemented as an overlay capability over an existing AMPS system. The details of HSCSD and CDPD are described in the following subsections.

6.1.1 High Speed Circuit-switched Data (HSCSD)

HSCSD is an enhancement of circuit-switched data (CSD), the original data transmission mechanism of the GSM. The objective of HSCSD is to provide a mixture of data and voice services without any changes to the physical infrastructure of current GSM networks. It gives the user one or more dedicated circuits for the entire call session and supports data-oriented application that were not available to the standard GSM users of the past. The examples of such applications are accessing LANs, e-mails, Internet on move, etc. The GSM CSD supports one user per channel per time slot whereas HSCSD allows a user to simultaneously access multiple channels (up to four) at any moment of time. This enables Internet access at the same speed of the many dial-up modem services across fixed line networks.

HSCSD architecture

The architecture of HSCSD service is similar to that of GSM. Since it operates across a GSM network, and therefore no additional hardware except an upgrade in network software is required to offer the service. It requires software upgrades in the MS and MSC (Mobile Switching Centre) with the help of which traffic stream is split into several streams and are combined again. The module called Inter Working Function (IWF) in MSC provides the required adaptation between the GSM network and the external networks. On the other hand, the Terminal Adaptation Function (TAF) module located at the mobile station (MS) achieves adaptation between the data terminals and the MS radio port. Both the TAF and IWF (Figure 6.1) perform splitting and combining of multiple composite data streams. In the A interface, the composite data streams are multiplexed into one 64 kbps circuit. In GSM, only the MS has voice capability. In HSCSD the MS is either the GSM MS that can be connected with a portable computer or a portable device with a built-in GSM MS.

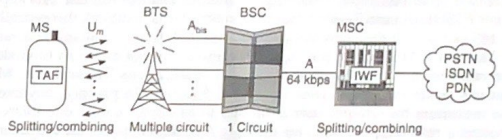


Figure 6.1 HSCSD architecture.

How does it work?

The parameters that characterize an HSCSD connection are used by the network to assign the network resources to a connection. The parameters are the desired number of channels or DNC, minimum or required number of channels or RNC needed for realization of the selected service, type of remote modem, etc. A user can choose the values of said parameters interactively during connection set-up and it can modify the values during the connection. As far as radio resource assignment is concerned, it is desirable that the number of traffic channels allocated in the forward direction, that is, from BSC to MS, is higher than that in the reverse direction.

The HSCSD supports two types of configurations—symmetric and asymmetric. In asymmetric configuration, more traffic channels are allocated in the forward direction. On the other hand, in symmetric configuration same number of traffic channels are allocated in both the forward and reverse directions. The network is configured to give priority to voice calls over data requests. That is, whenever there is a voice call to enable, it may not be possible for the system to satisfy a data request by providing all the requested slots.

The HSCSD extends the basic GSM circuit data service to higher speeds enabling allocation of multiple full rate traffic channels (TCH/F) to a single connection. In case of the symmetric configuration, if a data service is requested and multiple channels are available, the requested data is split into multiple streams by TAF in MS and is transmitted over U_m and A_{bis} respectively. Again these are combined by the IWF module in MSC and transmitted over the A interface. Similarly, when the requested data is served from external network through MSC, it is received over the A interface, split into multiple streams by IWF and is transmitted over A_{bis} and U_m respectively. To make the served data available to the user, the split data are further combined by the TAF module.

The error correction technique in HSCSD is different from that used in the original GSM service. In GSM, the significant part of transmission capacity is used for the error correction whereas HSCSD provides different levels of error correction and is used according to the quality of radio link. Following this technique, the existing 9.6 kbps data rate of GSM can be increased up to 14.4 kbps. Assuming GSM traffic channel full rate 14.4 kbps as per ETSI (European Telecommunication Standards Institute), theoretically an MS can use all the eight slots within a TDMA frame to attain data rate of $14.4 \times 8 = 115.2$ kbps. But in GSM, there is always a gap in time slot between forward and reverse direction communications. Therefore, an MS can use maximum 4 time slots in reverse direction. In practice, however, the maximum bit rate per user is limited to 64 kbps, as only one ISDN-B channel is reserved per user in the interface A of GSM network infrastructure.

Merits/demerits and applications

The HSCSD is acceptable for its theoretical simplicity. However, as it is using circuit-switched mechanism of GSM, the HSCSD-based system has to face the demerits of using circuit-switch instead of packet-switch. For n channels, the HSCSD requires n times signalling during a handover. In this case, instead of checking resources for a single channel, the BSC has to check resources for n channels. Therefore, the probability of blocking or service degradation increases in HSCSD during the handover.

As multiple channels are used, HSCSD charges higher rate from the subscribers than for a voice connection for the entire period. The subscribers are charged not on the basis of a voice call but on the basis of number of time slots allotted. It is cost effective for the steady traffic flow type of applications such as file downloading, video conferencing, etc.

6.1.2 Cellular Digital Packet Data (CDPD)

The Advanced Mobile Phone System (AMPS) offers data service using the CDPD overlay network which provides a 19.2 kbps data rate. The CDPD exploits idle periods on regular voice channels to provide the data service.

The CDPD system architecture contains Mobile Database Station (MDBS) in addition to the existing components at the cell sites. Figure 6.2 describes a CDPD network architecture.

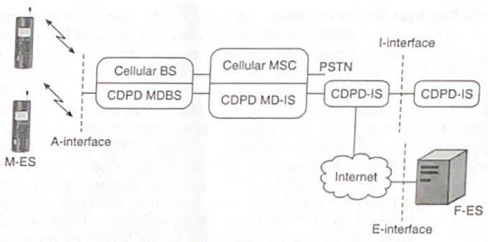


Figure 6.2 CDPD network architecture.

The major components of the CDPD network are:

- BS—Base Station
- MDBS—Mobile Database Station
- MSC—Mobile Switching Centre
- MD-IS—Mobile Data Intermediate System
- F-ES—Fixed END System
- M-ES—Mobile END System
- CDPD-IS—CDPD Intermediate System

Mobile End System (M-ES): An M-ES is the mobile host equipped with additional software and hardware to work in CDPD network. The software includes air-link protocol (shown in Figure 6.3) software, application software, etc. The air-link protocol is set for radio resource management, error-free transmission, and secured processing of data. The characteristics and complexity of application software differ with application. The example CDPD applications are the Internet access, periodic broadcasting of new information (if any) of an organization to its staff. The CDPD M-ES additional hardware comprises of chip set including DSP (digital signal processing) chip, microprocessor chip for control, etc.

The protocols at the network layer and above are contained in the Mobile Application Subsystem (MAS) module of M-ES. The *lower layers* protocols such as physical layer and data link layers are contained in

Subscriber Unit (SU) module of M-ES. The SU is responsible for establishing and maintaining data communications with the CDPD network. Each M-ES also contains a SIM (Subscriber Identity Module) to store the profile of the subscriber. Information stored in the SIM authenticates a subscriber.

The Internet or PSTN connection for an M-ES is established via MDBS, MD-IS and CDPD-IS. The MD-IS and CDPD-IS act as routers suitable for routing IP datagrams. When an M-ES does not transmit data, it goes to sleep mode for keeping the power down. It periodically keeps the power up to check for buffered incoming data.

Mobile Database Station (MDBS): All sorts of radio resource management that is, radio channel allocation, use of channels between CDPD and voice calls, keeping track of busy/idle status of reverse channels are done by the MDBS. The busy/idle status identification is important in CDPD, as it uses the idle periods on voice channels for data communication. The task of radio resource management is to ensure that the optimum radio channel is used by an M-ES. The optimality is evaluated in terms of reliability in performance while not interfering with other CDPD users or cellular users. Non-interference with cellular voice customers is of great importance, since CDPD is a service overlaid on the existing cellular voice network. The MDBS performs the tasks up to data link layer. It conveys data link information between an M-ES and the serving MDBS and is connected to the MD-IS. All of the MDBSs are connected to the MD-IS by microwave or wire-line links.

Mobile Data Intermediate System (MD-IS): In a cellular system, designed for voice services, the mobility management is taken care of by MSC/VLR (Visiting Location Register). Similarly, in CDPD system, the mobility management including roaming, billing and accounting support is taken care of by the MD-IS. The units that perform functions of the HLR (Home Location Register) and VLR of cellular voice network (Section 2.8) are the Mobile Home Function (MHF) and Mobile Serving Function (MSF) respectively in MD-IS. The MD-IS connects the external networks.

CDPD protocol architecture

The protocol stack at air interface has four layers, namely physical layer, data link and MAC layer, network layer and transport layer (Figure 6.3). At the physical layer, the M-ES uses Gaussian Minimum Shift Keying (GMSK), a digital radio modulation scheme. The GMSK is a variant of Frequency Shift Keying (FSK) modulation (Section 1.2.4). The bit stream over the physical interface is modulated using GMSK at data rate 19.2 kbps. The data is encoded to avoid errors during transmission.

The task of MAC (Medium Access Control) layer, a sublayer of data link layer, is to support channel sharing by multiple M-ESs while connecting the local MDBS. DSMA/CD (Digital Sense Multiple Access/Collision

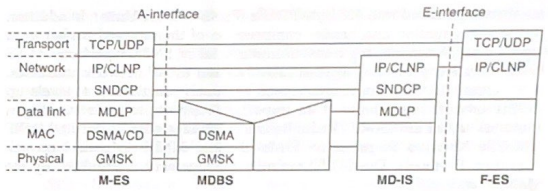


Figure 6.3 CDPD protocol stack.

Detection), a CSMA/CD[1] (Carrier Sense Multiple Access/Collision Detection) like multiple access technique, is used at the MAC layer. Since M-ESs cannot detect each other's transmissions, M-ESs are not treated as peers and thereby CSMA/CD can't be implemented in CDPD. In DSMA/CD, the MDBS takes care of the sharing of the channel so that no conflict arises out of accessing the network by the M-ESs. When an M-ES desires to transmit a data, it senses the busy/idle flag in the forward/downlink channel. If the channel is busy, it waits for a random period before sensing the channel again. This is referred as the defer mode. If it senses that the channel is idle, it initiates transmission and continues until it gets an indication from the forward channel to stop transmission. As soon as the transmission is stopped, it waits for a random interval of time and again senses the channel to check the busy/idle status.

The rest of the data link layer is controlled by MDLP (Mobile Data Link Protocol). The MDLP provides logical link control services between a Mobile End Systems (M-ES) and Mobile Data Intermediate Systems (MD-IS). It also provides basic error detection and recovery procedures.

The MDLP is a modified version of ISDN LAPD (Link Access Procedure on the D channel, Section 3.3.2). The modification is done to conserve bandwidth. Some of the procedures followed for conserving radio bandwidth are the selective rejection for retransmission of lost packets, multicast and broadcast addressing for the unacknowledged data service. In selective rejection mechanism, unlike LAPD, the receiver transmits a SREJ (Selective Rejection) frame for each missing frame. On receiving the SREJ, the sender retransmits only the frame identified by the SREJ frame.

The network layer runs on standard IP (Internet Protocol) and CLNP (Connectionless Network Protocol) suits. The SNDCP (Sub-network Dependent Convergence Protocol) compresses the unusually long network layer packet headers used in TCP/IP and CLNP. The compressed packets

[1] It is a protocol used in computer networking where a node detects collision before starting transmission by listening to the medium of transmission. The collision is detected if another node transmits at the same time. The desiring node stops transmitting and waits for a random time before trying to send the data again.

are then segmented into 128 byte PDU[2]s (Protocol Data Units). In addition to the segmentation and header compression of the received packet from upper layer, for increasing transmission efficiency, SNDCP performs M-ES authentication protecting against unauthorized use of network addresses.

Other than the above mentioned protocols functioning at levels up to the network layer, some of the important support protocols working at network layers are RRMP (Radio Resource Management Protocol), MNRP (Mobile Network Registration Protocol), and MNLP (Mobile Network Location Protocol). The CDPD mobility management is provided by the MNRP and MNLP.

The RRMP is a network layer protocol to manage the available radio resources for efficient operation of the CDPD system. The RRMP at M-ES uses the services of physical layer and data link layer to perform functions such as selection of the best radio frequency channel for setting up and maintaining its connection to the network without interfering with the existing voice service. The task of RRMP at M-ES is also to help to perform an M-ES controlled handoff. Each MDBS sends PDUs containing parameters such as received signal strength, block error rate, etc. regularly to all the M-ESs in the locality. With the help of these parameters an M-ES can control handoffs.

The MNRP operates over air interface and is used by the M-ES to declare its presence to the serving system. When an M-ES desires to register to a system, it selects the CDPD channel with the best signal strength and informs it to the corresponding MDBS. The serving MD-IS updates the mobility manager inside it so that after this all data for the M-ES can be routed to that cell. The MD-IS also uses this protocol to inform the intended M-ES to provide its service. In short, the MNRP performs the registration/deregistration and authentication of an M-ES and confirmation by an MD-IS of its readiness to provide services to an M-ES.

The MNLP resides at the network layer of MD-IS and operates in conjunction with MNRP to provide mobility management. In case of a roaming, the serving MD-IS notifies the home MD-IS using MNLP that it is providing service to the M-ES. It exchanges location information of an M-ES between MD-ISs for forwarding and routing of messages to the roaming M-ES. In a word, MNLP notifies to MD-ISs about a registered M-ES's authentication, confirms about the readiness of the MD-IDs to forward packets to an M-ES and forwards packets from the home MD-IS to the serving MD-IS.

Channel assignment and hopping

There are mainly two channel assignment strategies by means of which the radio channels may be configured and assigned in a CDPD system

[2] In a layered protocol architecture, information that is sent/received as a unit containing address, control information and data among peer entities of a network is called PDU. For example, the PDU in physical layer is bit whereas the PDU in network layer is packet.

(overlaid on an existing AMPS system). The strategies are fixed assignment and shared assignment.

Fixed assignment: In this strategy, a predetermined number of channels are assigned to a CDPD system from the AMPS pool. An assignment is permanent and avoids sharing of channels between voice and data parts of the system. That is why it is the simplest design and also of low cost. However, as it defeats the main purpose of CDPD that is, the sharing of the unutilized capacity of cellular (AMPS) service, it can't be so effective in practical design.

Shared assignment: In such a case, the channels are fully or partially shared between voice and data traffic. A set S_c of channels is declared accessible to CDPD in a partial sharing scheme. The mechanism is to assign voice calls from the channels (if available) dedicated for the AMPS system. If there is no such channel available, satisfy a voice call from the set S_c of shared channels. In full-sharing scheme, the complete set of AMPS channels belong to the S_c.

Channel Hopping is the concept of using unused radio resources between the voice calls. This attempts to create a radio frequency data channel out of apparently unused and unusable radio resources. The queuing analysis of telephony traffic models point to the fact that for an acceptable call blocking rate, there can be a considerable number of channels which becomes surplus even in the busy hour. The CDPD implements packet-switched data service exploiting this surplus channel capacity. It avoids conflict with cellular (AMPS) voice traffic employing a mechanism referred to as sniffing and channel hopping. The MDBS sniffs radio signal from the transmission path and analyses it for voice activity. For identifying the unused voice channel, the sniffer, a module in MDBS, scans and analyses each of the AMPS 30 kHz channels. There are mainly two types of channel hopping–forced/emergency hopping and timed/planned hopping.

A planned hopping means the MDBS periodically switches the CDPD data traffic to different channels. This demands an intelligent MDBS that should learn the channel assignment algorithms of cellular system followed by different manufacturers. Further, there may be instances where a voice call is assigned to a channel occupied by CDPD data traffic. The issue, however, is addressed through the use of sniffer in MDBS. The sniffer senses the initiation of a voice call and continuously monitors the non-CDPD (voice) transmissions. If such a transmission is detected, the MDBS terminates the CDPD (data) transmission and reassigns the CDPD traffic to an unoccupied channel.

The priority of voice traffic is higher than the data traffic. While a voice call arrives at the AMPS BS, it selects an idle channel. If the channel is busy with a CDPD data call, the data transmission is suspended until another idle channel is found. This is true for both the fixed and shared channel assignment strategies. The MDBS reestablishes the suspended data transmission switching to another channel it finds free.

6.2 GENERAL PACKET RADIO SERVICE (GPRS)

To compete with the AMPS-based CDPD system supporting packet-switch mode of data service, a GSM-based data service, GPRS (General Packet Radio Service) was evolved as a successor of HSCSD. Primarily GSM was designed for circuit-switched services and was then upgraded in order to support high-speed packet-switched services. The GPRS system uses a packet-switched connection on the radio interface.

The objective of enhancing GSM, led by ETSI, was to specify the service that accommodated data connections with high bandwidth efficiency. However, the main intention of GPRS is to enhance the range of existing GSM data services. GSM offers data rates up to 9.6 kbps whereas GPRS data rate is up to nearly 170 kbps. It can be flexibly allocated according to the actual user demands.

The offered data rate depends on the network availability, channel coding scheme and terminal capability. The increased speed in GPRS with respect to GSM is achieved by using more than one timeslot (maximum 8 time slots) of the TDMA frame. The packet-switched characteristic enables varying allocation of the available timeslots in different instant. For example, a system may have 8 timeslots at time t_1 and say 4 at t_2. If the GPRS system sacrifices GSM's error correction capabilities, it can have 21.2 kbps data rate. If 8 time slots are allocated, the system can achieve $8 \times 21.2 = 169.6 \approx 170$ kbps.

The GPRS reserves radio resources only when there is data to send and reduces reliance on traditional circuit-switched network elements. The increased functionality of GPRS increases the penetration of data services among consumers and business users and as a result incremental cost decreases to provide such services to the users.

6.2.1 GPRS Architecture

The GPRS is an enhancement of GSM. It adds some nodes in the network to provide the packet-switched services. These network nodes are the GPRS Support Nodes (GSNs) and are responsible for the routing and delivery of data packets. The data is routed from the mobile station to external packet data networks such as Internet or X.25 and vice versa.

Figure 6.4 describes the GPRS network architecture. Two GSNs, the Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN), are added with the existing GSM network. A hardware module called packet control unit (PCU) is included in the BSC. The PCU is responsible for providing 'capacity on demand' feature. It dynamically selects radio resources to be allocated for packet-switch or circuit-switch use. The PCU manages radio resources for GPRS traffic whereas a BSC manages radio resources allocated for circuit-switch use. In addition to enhancement of BSS module, a software upgrade is done in the existing BTS and BSC. The SGSN is connected to the BSS via a frame relay connection to the PCU in the BSC. The GGSNs connect external data networks such as Internet, X.25.

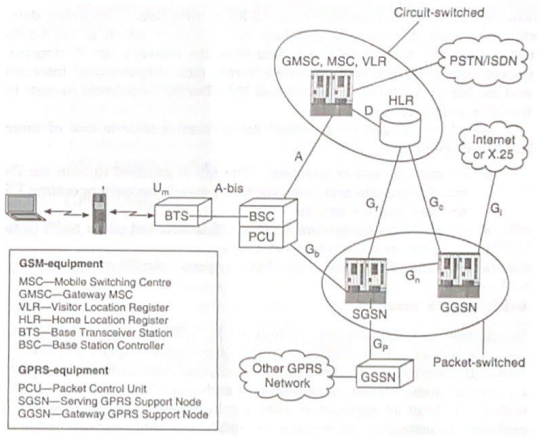


Figure 6.4 GPRS architecture.

As the existing GSM terminals can neither handle the enhanced air interface nor packetise data directly, a mobile station accessing the GPRS network should be the GPRS-enabled phones, PDA, etc. These stations are backward compatible with GSM for voice calls. Therefore, to use GPRS, users specifically need a mobile phone or a mobile station that supports GPRS protocol stack and subscription to a mobile telephone network that supports GPRS protocol stack. The database, including VLR and HLR database, also require software upgrades to handle the new call models and functions introduced by GPRS.

GPRS operates in both circuit-switch (CS) and packet-switch (PS) domains. In CS domain an MS is identified by IMSI (International Mobile Subscriber Identity) and TMSI (Temporary Mobile Subscriber Identity) as in GSM (Section 3.3.1). In PS domain, an MS is identified by IMSI and P_TMSI (packet TMSI). Both the temporary identities (TMSI and P_TMSI) are for concealing user's identity IMSI. While in GSM the current VLR of an MS assigns TMSI, in GPRS the SGSN of an MS assigns P_TMSI.

Prior to accessing the GPRS network, an MS registers to the SGSN. The SGSN in turn validates the user consulting the user data stored in HLR. Upon validation, the SGSN assigns a P_TMSI to the user. This process of connection to the GPRS network is known as GPRS attach. Similarly the disconnection from the network is called GPRS detach. A node can get

different services such as SMS over GPRS, notification of incoming data, etc. only when the node is attached. An MS must attach to the GPRS network before transferring any data over the network. In PS domain, packet data is transferred between external data network (e.g. Internet) and the MS through GGSN, SGSN and BSS. The SGSN delivers packets to the MSs within its service area.

Based on the service domains, an MS can operate in one of three modes of operation:

- *PS and CS mode of operation:* The MS is attached to both the PS and CS domain and it is capable of simultaneously operating PS services and CS services.
- *PS or CS mode of operation:* The MS is attached either to PS or to CS domain but only one at a time.
- *PS mode of operation:* The MS supports only PS connection.

6.2.2 GPRS Services

In addition to GSM services (Section 3.3.3), GPRS provides multimedia message service (MMS), SMS with higher transmission speed and Instant messaging service. MMS allows sending messages with multimedia objects, e.g. images, audio. Instant messaging is the exchange of messages in real-time with the help of an application software. In instant messaging, message exchange is instant in comparison to e-mail, and continued exchange is simpler than sending e-mail to and fro.

GPRS supports Internet access through wireless application protocol (WAP). It also allows broadcast, multicast and unicast services.

Traffic types supported in GPRS are categorized as:

- conversational class (video conferencing)
- streaming class (one-way video)
- interactive class (web browsing, database access)
- background class (e-mail, SMS)

In the conversational class services, communication is bi-directional and real time. The communication in streaming class services are also real time but unidirectional. Both the interactive and background classes are non-real time communication. Different traffic types need different QoS. Users can negotiate the QoS requirements. Examples of QoS parameters are traffic class (e.g. interactive), maximum bit rate (e.g. 128 kbps), transfer delay (e.g. best effort), etc.

From the user's point of view GPRS is more convenient for the following reasons:

- Enhanced speed (nearly 170 kb/s)
- new services provided
- unlike time-oriented charging applied for circuit-switched connections, packet-switched data service in GPRS allows charging depending on the amount of data transmitted and the quality of service negotiated

The application areas where the GPRS can provide attractive mobile data communication services are:

- Communications—Internet access, e-mail, fax
- Bandwidth on-demand for point-to-point transmission
- Value-added services—stock prices, games
- E-commerce—banking
- Multicast and group call service
- Location-based services—hotel and restaurants in the locality of MS's current position

6.2.3 GPRS Channels

Like GSM, the GPRS also uses physical and logical channels to carry the data and call controlling signals. The only physical channel called PDCH (Packet Data Channel) is used in GPRS. However, the GPRS uses a number of logical channels that are mapped into physical channels. There are four types of logical channels—dedicated traffic channel, common control channel, dedicated control channel and broadcast channel. For initial set-up, the GPRS uses GSM control and broadcast channels (GSM logical channels). The other tasks are performed utilizing other logical channels (Refer to Figure 6.5).

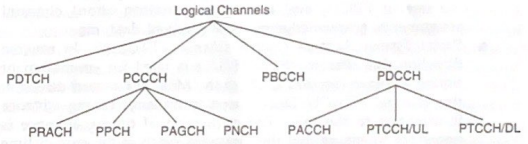


Figure 6.5 GPRS channels.

Dedicated traffic channel

It is known as Packet Data Traffic Channel (PDTCH) and is used for user data traffic transfer. It is a bi-directional channel. A PDTCH is temporarily dedicated to a user or group of users. For forward and reverse directions, the PDTCHs are assigned separately to support asymmetric user traffic flow. One time slot in a TDMA frame can be shared by a maximum of 8 PDTCH channels that is, 8 mobile stations can share one TDMA time slot for transmitting data packet.

Common control channels

It is also called Packet Common Control Channel (PCCCH). The common control channels are utilized for signalling the packet data. The following are the common control channels in GPRS.

- *Packet Random Access Channel (PRACH):* This random access channel acts in a reverse direction. An MS starts sending burst of data towards the base station using PRACH.
- *Packet Paging Channel (PPCH):* This is the forward channel used to make the MS aware of the incoming call. Alternatively, the PPCH is used for signalling control prior to a call set-up. Once the call is started, a dedicated control channels PACCH gets activated.
- *Packet Access Grant Channel (PAGCH):* It is a forward channel. It informs the MS about the traffic channel assigned to a call.
- *Packet Notification Channel (PNCH):* This control channel also acts in a forward direction. It alerts a group of MSs that there is a broadcast/multicast traffic targeted for the MSs.

Dedicated control channels

Following are the two channels of Packet-Dedicated Control Channel (PDCCH) type. These two channels act in both the forward and reverse directions.

- *Packet Associated Control Channel (PACCH):* While a call is in progress it is used to carry signalling information to and from the MS. The PACCH is utilized once a call is set-up (completion of the use of PPCH) and carries information about channel assignments, acknowledgements of received data, etc.
- *Packet Timing Advance Control Channel (PTCCH):* In reverse direction this channel (PTCCH/UL) is used for estimation of timing advance (Section 2.7.2) of the MSs. In forward direction this channel (PTCCH/DL) is used to transmit timing advance information to the MSs. The estimation of timing advance is necessary to ensure that the messages reach at the correct time at base station irrespective of the distance between the MS and the base station.

Broadcast channel

The Packet Broadcast Control Channel (PBCCH) is the forward channel and is used to broadcast system information such as power control parameters, network parameters, etc. This information is required to the MSs within a cell for setting up a call or for transferring data. The PBCCH is comparable to the BCCH of GSM. During the initial set-up, the GSM BCCH is used to provide a time slot number for PBCCH.

6.2.4 GPRS Protocol Stack

The protocol stack for GPRS is specially designed to ensure seamless access to external data networks such as Internet and PSPDNs. Figure 6.6 is the GPRS protocol stack describing the message flow from MS to GGSN.

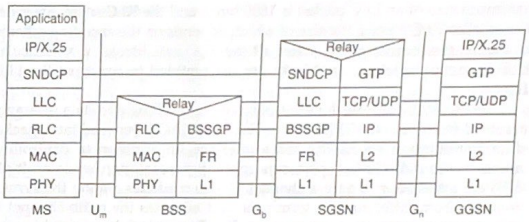


Figure 6.6 GPRS protocol stack.

The network protocol for GPRS backbone is IP. The Base Station Subsystem GPRS Protocol (BSSGP) is used to convey routing and QoS-related information between the BSS and SGSN. BSSGP works on top of a Frame Relay (FR) network, however, it does not perform any error correction.

The Subnetwork Dependent Convergence Protocol (SNDCP) in association with logical link control (LLC) is used between an MS and SGSN to adapt with the different characteristics of underlying network components. The SNDCP provides the functionality to map different network protocols onto the logical link control (LLC). This includes multiplexing of packets from different protocols, header compression, data compression, data encryption, and segmentation of packets. Data compression is done to reduce load on the radio channel and segmentation is needed to match the LLC packet size. SNDCP encapsulates the IP packets in GPRS specific packet formats.

LLC, the higher sub-layer of the data link layer, is primarily dedicated for establishing a logical link between an MS and the SGSN. It also provides a safe logical link introducing the encryption of packets. This logical link is independent of the underlying radio interface protocols. The LLC protocol is based on LAPD (Section 3.3.2) used in GSM and can support point-to-multipoint (PTM) transmissions. The data transmission can either be acknowledged or unacknowledged. The acknowledged transmission provides selective retransmissions of data packets that have not been properly decoded at the receiving end. The acknowledged mode of transmission is for error-free delivery of data packets. On the other hand, in unacknowledged mode of transmission, retransmission is ignored. This mode is suitable for the applications that are tolerant of error such as video streaming. When an MS moves to a new routing area under the different SGSN, the current LLC is removed and a link is established with the new SGSN.

RLC and MAC layers provide a reliable radio link for data transfer between an MS and the BSS. The radio link protocol performs error correction. The RLC segments an LLC frame into RLC data packets. The

maximum size of an LLC packet is 1600 bytes and the RLC segments them into smaller RLC blocks, the size of which depends on the encoding scheme in use. At the receiving end, when all the RLC data blocks of a particular LLC frame are reached, the blocks are reassembled to construct the LLC frame.

MAC layer controls the multiplexing of signalling and data messages received from various GPRS users. The BSS MAC layer facilitates packet data transmission across the radio interface. In addition to controlling access to the radio link, it performs contention resolution while multiple MSs are attempting to access channel. It also mediates among the service requests from different data terminals, and allocates the radio channel to individual data terminals on request. The RLC/MAC layer combination ensures flexible allocation and utilization of radio resources in the GPRS network.

The physical (PHY) layer controls the physical channel management. The tasks include modulation, demodulation, transmission, power control, and channel coding/decoding. This layer also performs error correction and interleaving. The physical channel can be encoded by means of one of the standard available channel coding (Section 1.2.4) methods. The choice is based on a trade-off between bit-error rate and throughput across the channel.

All data within the GPRS backbone, i.e. between SGSN and GGSN are transferred following the GPRS Tunnelling Protocol (GTP). It uses G_n interface. As a result, the internal backbone network does not have to deal with IP addresses outside the GPRS network. The GTP can use either the reliable TCP (needed for reliable transfer of X.25 packets) or the non-reliable UDP (used for IP packets).

6.2.5 Mobility Management and Data Routing

One of the main issues in GPRS network is the routing of data packet to and from mobile station and external packet data network. Compared to the GSM's two states (dedicated, idle) model, in GPRS an MS can have three states—ready, stand-by and idle. To act as GPRS node, an MS keeps its switch on and sets up a PDP (Packet Data Protocol) context with the network. The context set up means the node is attached to the GPRS network, that is a logical link is established between the MS and the SGSN. The context information stored both in the MS and in the corresponding SGSN. As a result, the MS changes its state from idle to ready. If for a predetermined time period, no data transfer takes place to and from the node, it enters into a stand-by state. An MS can transmit data only at a ready state. So if there is any such data for transferring, the MS goes back to the ready state. Further, from the ready state if GPRS detach is performed, it can go to the idle state and all the PDP contexts are destroyed. If a predetermined time expires, the MS goes to the idle state from the stand-by state.

Mobility management

In ready state, the SGSN knows cell location of the MS. However, in stand-by state, the location identified is the routing area. The routing area is defined as a group of adjacent cells. The idea of stand-by state reduces the load in GPRS network caused by the cell-based routing update messages, and it also conserves the battery power of an MS. A cell-based routing update procedure is invoked when an MS, in ready state, enters a new cell. When an MS in a ready or a stand-by state and moves from one routing area to another in the service area of same SGSN, it must further perform a routing update. The routing area information in the SGSN is updated.

In the idle state, an MS neither has any logical GPRS activation nor any Packet-Switched Public Data Network (PSPDN) addresses allocated to it. In this state the network does not even know the MS's location.

Routing

The gateway GSN (GGSN) interacts with the external data network and routes external data packets to the SGSN. Before data can be sent or received, the GGSN must create a PDP context for the subscriber. A PDP context data is stored in MS and in SGSN. The information includes MS's status (idle/ready/stand-by), P_TMSI, routing area, cell_id, QoS, etc. The PDP context assigns a PDP address (e.g. IP address) for the communication and associates it with the IMSI, and the address of the SGSN. A PDP context can identify the application that is being used (e.g. web browsing, streaming video) and it can also specify the requested QoS. A subscriber can have several PDP contexts activated at one time.

Data is transmitted between an MS and the GPRS network only when the MS is in ready state. When the SGSN sends a packet to an MS that is in the stand-by state, the MS must be paged. Because the SGSN knows the RA in which the MS is located, a packet paging message is sent to that RA. After receiving the packet paging message, the MS gives its cell location to the SGSN to establish the ready state. In the idle state it is not possible to send messages to the MS from external data networks such as PSPDN. The MS can receive only multicast messages that can be received by any GPRS protocol stack.

MS-originated PDP context creation

If an MS desires to create PDP context, it sends a request to the SGSN. The request message (PDP context) includes parameters such as IP address of the MS, requested QoS, etc. The SGSN first validates the MS. It then derives the GGSN address from the parameters sent along with the PDP context. Once the GGSN address is found, the connection will be established between the MS and the external data networks through the GGSN. The SGSN creates a downlink GTP tunnel for routing of data packets from GGSN to SGSN. On the other hand, the GGSN creates an uplink GTP

tunnel from SGSN to GGSN. The GGSN then responds the request from SGSN by sending PDP context creation information along with negotiated QoS. The SGSN in turn responds the PDP context creation request from the MS and sends the negotiated QoS.

Network-originated PDP context creation

When a data packet reaches to the GPRS network for delivery to an MS, the GGSN decides to establish a PDP context if no such context has already been established. With the help of the MS identifier IMSI, consulting HLR, the GGSN finds out the SGSN of the MS. The GGSN sends a request to the SGSN to take an action for the packet delivery. The SGSN in turn forwards the request to the MS. The MS then originates PDP context creation as it does in case of 'MS-originated PDP context creation'.

Similar to PDP context creation, PDP context deletion may be originated by an MS or the GPRS network.

6.2.6 GPRS User Validation

In GPRS, the user validation process is similar to the user validation in GSM (Section 3.3.8). When an MS performs attach process, a 3-tuple (random number, signed response, authentication key) is created and stored in SGSN. During the attach process, the SGSN also sends the P_TMSI (Section 6.2.1) to the MS. The MS inserts the P_TMSI in the subsequent routing area update request sent to the SGSN. If the P_TMSI does not match, the SGSN starts user validation process.

An MS contains IMSI, authentication key, algorithm, ciphering key. The SGSN sends the random number to the MS as challenge. The MS feeds the random number and authentication key to the algorithm to generate a signed response. The computed signed response is then sent to the network. The network compares the received signed response with the signed response stored in the SGSN. If there is a match the user/subscriber is recognized as valid.

Similar to GSM, upon confirming the user-validity ciphering key is generated at both the ends of the air interface and the key is used for encryption and decryption of data. There is one major difference between GSM and GPRS ciphering. In GSM, encryption/decryption of data is performed in the air interface between MS and BSS. On the contrary, in GPRS it is performed in the air interface between MS and SGSN.

6.3 WIRELESS APPLICATION PROTOCOL (WAP)

Next to GPRS, Wireless Application Protocol (WAP) popularizes wireless Internet services over the cellular network. At the initial phase of development of data service, the HSCSD, CDPD and GPRS targeted increase in data rate. In spite of directed efforts, the enhanced data rate could be attained in the order of kbps. The Internet designed for fixed

network was not designed to run in this low bandwidth connection. Even workstations over fixed network run TCP/IP in Mbps range. Further, the Internet protocols, i.e. TCP/IP and HTTP are not suitable for use in mobile phone communications. These introduce too much overheads requiring many messages between client and server simply to set up a connection, thus requiring a high processing power at the client device. Moreover, the information flow from the Internet to the mobile phone passes through various elements in the wireless network, each one introducing a delay. Therefore, as the Internet protocols transmit large number of voluminous messages, it naturally results in a significant delay. These limiting factors of Internet protocols initiated the search for a new set of protocols appropriate to access the Internet with portable devices in wireless network.

Internet access from client in wireless network

The large bandwidth requirement is essential for Internet access as it demands transmission of huge data due to large header content from server, repetition of some information with each request and transfer of data without compression. One solution of successful Internet access from wireless network is increase of data rate. The alternative can be the introduction of a mechanism so that the Internet access needs less volume of data transfer from server to client requiring lower bandwidth. One way of implementation of such mechanism is trimming of the content before it reaches the client from the server. This can be achieved by scaling down the contents. The simplest examples are: conversion of colour images to monochrome, curtailment of advertisement, availing the entire content in a selective manner, etc. Many of these techniques are implemented by replacing the existing Internet language (HTML) with WML (Wireless Markup Language). The alternative practice of reducing data transfer is the conversion of a document page (e.g. pdf to txt) in a form that avoids loading of a special software (e.g. acrobat reader) for presentation. The pushing of some of the frequently accessed information (road condition, score of a match, stock exchange news, etc.) to all the clients from the servers also can reduce the volume of data transfer. WAP PUSH is a function that enables applications to send information to a WAP end-user with no previous user action required. PUSH messages also open up possibilities for new applications where end-users can receive information of interest.

WAP has become a synonym of new, wireless Internet Services. The unmanageable overhead associated with TCP/IP to take care of all sources of data loss such as network congestion, transmission errors, handoff disruption is the main drawback in wireless environment. This probably motivates wireless equipment manufacturers and service providers to form a Forum to enable low-overhead Internet access from wireless application devices with limited computation power and constrained battery life. In

1997, the major cellular phone manufacturers, e.g. Ericsson, Nokia, Motorola, etc. joined to form the WAP Forum. It is an industry group responsible for managing and extending the WAP standard and facilitating the adoption of WAP.

The main advantage of WAP over TCP is less overhead. In WAP-enabled environment, a website transmits scaled-down versions of normal web pages specifically optimized for the use by wireless telecommunication devices. WAP targets to provide a platform where access to varied Internet content and different data services available from diverse types of mobile terminals across heterogeneous types of wireless network such as GSM, GPRS, CDPD etc. is simplified.

WAP is a standard for transmission and presentation of information over a wireless connection. With WAP, the wireless service providers can provide high-level interactive information services to subscribers over the same voice network. A request-response procedure is used to establish the WAP connection. It can be modeled as a client server environment where MS is the client and the Internet host is the server.

6.3.1 WAP Architecture

The components of the WAP architecture contains a WAP-enabled device (mobile client), a public land mobile network such as GSM/GPRS, a public switched telephony network such as ISDN, a WAP gateway, an IP network and WAP application/web server. Figure 6.7 shows the components of WAP architecture and interconnection among the components for Internet access from a WAP-enabled device connected with a GSM/GPRS network. The diagram also describes the protocol stacks used in WAP-enabled device and in WAP gateway. The functionalities of the WAP gateway and the protocol stacks are elaborated below.

A WAP gateway is placed in between mobile networks and Internet, that is, between clients and server. It works as a protocol converter between WAP protocols and common HTTP/TCP. The WAP client communicates with the gateway via WAP protocols. The resources are fetched by the gateway from the web server using HTTP. A client submits its request in a coded form to get a response from the remotely located server. However, the target is to reduce the volume of data transfer over the air. Therefore, the encoding of request data to convert it in a compact form is very much essential. The decoder present in Gateway converts the request in a form understandable to the server (e.g. web server). The server treats the request as if submitted from a fixed node/client. While the web server responds, encoder at the gateway again converts (encode) it into WAP client understandable form. In a word, the gateway acts as a converter between the WAP protocol stack and the WWW stack (HTTP, TCP/IP). The encoded information is in a more compact form and that leads to reduced volume of data transfer over air. The web server provides content in the form of

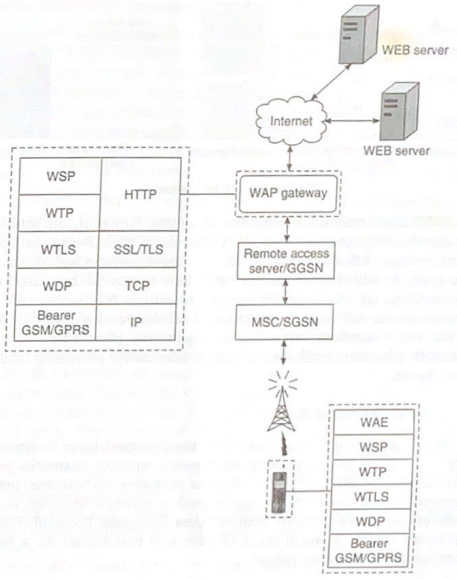


Figure 6.7 WAP architecture.

HTML-encoded pages that are transmitted using the standard web protocol stack (HTTP, TCP/IP). The HTML content is then filtered to translate it into WML content. Finally, the WML content is converted to binary WML and delivered to the client following WAP protocol stack.

In case of PUSH service, the WAP gateway (Figure 6.8) acts in the same manner as in case of client-initiated service request described in the previous section. Whenever a web server pushes content to the gateway, it encodes the pushed content and sends the same to the client. In fixed network environment, normally the fixed node/client caches frequently accessed content inside the client itself. On the contrary, in WAP, as the client has limitations in terms of storage and computation capacity, the gateway maintains the cache if required.

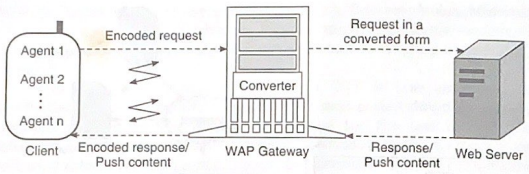


Figure 6.8 WAP gateway.

WAP client maintains a number of agents. However, the number of such agents is not specified in the WAP standard. Each client must contain at least one agent that supports WML. The agent helps in web access, data editing, etc. In addition to this, the other user agents can be placed to get other services, such as telecommunication services from within the same terminal, phone call with the facilities of call forwarding, call answering, etc. The WAP standard only specifies capabilities of user agents, e.g. OS version, display size. It allows vendors to offer newer services introducing newer agents.

6.3.2 WAP Protocol Stack

The WAP suite protocols operate from the transport layer to application layer. It is designed to operate over many wireless networks such as GSM/GPRS, HSCSD, CDPD, etc. Thus the protocols are (wireless) network-independent and work on the top of such a network. In other words, it works on the top of common Internet User Datagram Protocol (UDP/IP). The layers of the protocol stack (Figure 6.9) maintained by a terminal (client side) is described below.

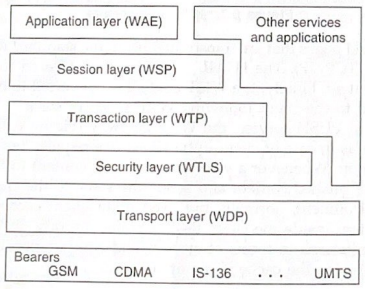


Figure 6.9 WAP protocol stack (client side).

Wireless Application Environment (WAE)

The application layer provides a network-neutral application environment and permits a high degree of device independence by using the wireless application environment (WAE). The WAE includes: (i) wireless markup language (WML) which accommodates limitations of the wireless devices towards display capabilities; (ii) WML-script which is WML's accompanying client-side scripting language provided for additional intelligence and control over presentation, and (iii) wireless telephony application which allows call control, network text messaging, and also offers a phonebook interface.

WML: Wireless markup language (WML) is the presentation layer for WAP on top of the WAP stack. It is XML-based presentation language for creating micro user interfaces, essentially equivalent to HTML for the PC. WML introduces new features suited for use with mobile devices but lacks many of HTML's features. The WML pages are called cards. Each card contains content and navigational controls. WML preserves content of variables of its cards/pages. Unlike HTML, WML is not stateless. In wireless network delay in connection to the web server is higher than in wired environment. Therefore, instead of transferring one page at a time (as in HTML) between the device and the server, a number of cards at a time is transferred reducing the delay. A group of cards transferred between the web/application server and WAP device in a single transmission unit is called deck. The cards and decks are defined in WML to reduce client/server interaction that effectively reduces the response time for user interactions. The following is a sample deck with single card:

```
< ? xml version = "1.0"?>
< ! DOCTYPE wml PUBLIC "- //WAPFORUM//DTD WML 1.1//EN"
http://www.wapforum.org/DTD/wml_1.1.xml>
<wml>
   <card id = "page1" title = "A Simple Example"
         <p>
               Welcome to WAP
         </p>
   </card>
</wml>
```


The first two lines are xml headers. Every deck starts with such headers. The wml code of a deck is enclosed within <wml> </wml> tag pair whereas the cards within the deck are enclosed <card></card> tag pair and the text to be displayed is enclosed within paragraph tag pair. WML has a fewer tags than HTML. When a WAP device receives the above card, it will display **Welcome to WAP**.

WML script: It is derived from Javascript. It provides programmatic extensions to WML user interfaces, much the same way Javascript can extend HTML web pages. WML contains the scripting language for performing simple tasks on the WAP terminal, such as validating user

input. It also avails basic programming functions (loops, conditional statements, etc.) optimized for minimal memory usage. The standard library functions available in WML script are:

- Base functions—max, min, etc.
- String manipulation functions—substring, length, etc.
- Floating point functions—floor, round, etc.
- Functions for manipulating the associated WML context–loading a new page, updating the display
- Dialog library

WML script file is stored as a separate file whereas javascript is embedded in the same HTML file.

As mentioned earlier WAP devices have limitations in terms of storage and computation capacity. It also has the problem of limited bandwidth. At the time of receiving an application service, if the device has to access the server even for a small computation, the entire service becomes slow. If WML script is used, the said computation/processing is done on the client and the delay caused by the limited bandwidth is thereby reduced. An example of WML deck with a call to WML script is shown below:

```
< ? xml version = "1.0"?>
< ! DOCTYPE wml PUBLIC "- //WAPFORUM//DTD WML 1.1//EN"
   http://www.wapforum.org/DTD/wml_1.1.xml>
   <wml>
   <card id = "page2" title = "Example calling script">
                  <p>
                        Computed value of "5+6" is : $ (summ) <br/>
                        < a href = "scrpt.wmls # addedresult (5,6)" >
                              Compute
                        </a>
                  </p>
            </card>
   </wml>
```

Here <a> tag pair is used to call the script file setting href attribute. The
 tag is used to give a line break, that is, a new line is started after the
 tag. Calling wml script is same as referencing another wml card by <a> tag pair. To refer to another wml card, the corresponding link is assigned to the href attribute. In the above example, the reference points to a wml script scrpt.wmls. Content of the scrpt.wmls is as follows:

```
extern function addedresult (x,y)
   {
         var v;
         v = x+y;
    WMLBrowser.setVar("summ", v);
    WMLBrowser.refresh ( );
   }
```


When a WAP device receives the wml deck with a call to the wml script, it computes the addition of 5 and 6 and refreshes the browser with the following result:

Computed value of "5+6" is : 11
Compute

Wireless Session Protocol (WSP)

WSP provides a steady interface to the application layer for two session services—the connection-oriented service and the connectionless service. The connection-oriented service operates on top of the transaction layer protocol. On the other hand, the connectionless service operates above a secure or non-secure datagram transport service/protocol. A wireless session protocol enables the WAP client to negotiate, open and maintain a session with the WAP gateway. It also supports HTTP 1.1 functionality and semantics in a binary encoded format to minimize data transfer to a WAP terminal.

Wireless Transport Protocol (WTP)

The WTP can be considered equivalent to the TCP layer in TCP/IP protocol suite. It is optimized (made light-weight) for low bandwidth but even then it ensures a reliable data transport mechanism. The most salient feature of WTP is that it relieves the application layer of the tasks of retransmissions and acknowledgements that are necessary for datagram services. WTP offers three classes of transaction services as follows:

Class 0—unreliable, one-way messages without confirmation of receipt of the messages, e.g. weather information (unreliable PUSH service).

Class 1—reliable one-way message without the result messages, e.g. SMS (reliable PUSH services) where no response is expected.

Class 2— reliable two-way request-response messages, e.g. every request is answered with a response.

Wireless Transaction Layer Security (WTLS)

This layer is designed to provide cryptographic services for privacy, data integrity and authentication between two communicating applications. It also avails an interface for creating and terminating secure connections and provides the upper sublayer of WAP with a secure transport service interface.

Prior to a data exchange in WAP the WTLS session has to be established by performing an initial key exchange and negotiation of the cryptographic algorithm. The privacy of communication can be secured using the Data Encryption Standard (DES) or the International Data Encryption Algorithm (IDEA).

Wireless Datagram Protocol (WDP)

The WDP in WAP is a general datagram service. It offers a consistent service to the upper layer protocols and communicates transparently over

one of the available underlying bearer services. This enables upper layers to operate independent of the underlying bearer services by providing a uniform interface. The errors during transmission at the WDP layer is communicated via a wireless control message protocol (WCMP), similar to that of Internet control message protocol in the IP world. The WAP interface also provides an independent control channel for error reporting and administration.

Bearers

The WAP protocol stack (Figure 6.9) is designed to operate with a variety of bearer services. HSCSD, GPRS and UMTS are some of the bearer services in WAP. HSCSD is intended for the user who wants dedicated bandwidth, e.g. immediate file transfer, while the GPRS suits for the users who don't require such time-sensitive services. The HSCSD mainly differs from the other technology platforms in terms of its ability to support high-speed real time data. However, packet-switched bearer services (e.g. GPRS) are much better suited than circuit-switched (e.g. HSCSD) services for mobile devices. The packet-switched bearer services can provide comparatively more reliable service in an inherently unreliable mobile environment.

WAP gateway protocol stack

The WAP gateway maintains two protocol stacks, one for peer communications with the WAP terminals through wireless link, and the other for peer communications with the Internet server through fixed link. At one end, the gateway provides interface with the WAP protocol stack over wireless link and at the other end, it offers interface with Internet protocol stack over the fixed link. Figure 6.10 shows layers of protocol stack at WAP gateway along with the corresponding layers at two of its peers (WAP terminal and Web server).

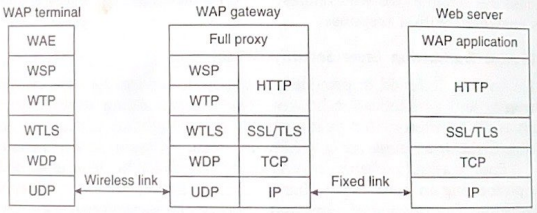


Figure 6.10 Protocol stack (WAP gateway and peers).

The following are the steps of operations while accessing the Internet from a WAP terminal.

- From a WAP terminal, the user submits a binary encoded request filling up a WML page to a web server.

- The WAP gateway intercepts the request, decodes it and translates it into a standard HTTP request and then forwards it to the desired web/application server specified in the URL given with the request.
- The web/application server receives and processes the request.
- On receiving the request, the web server responds in a standard HTTP format to the WAP gateway.
- The WAP gateway encodes the response and forwards it to the WAP terminal.
- The WAP terminal receives the response message and renders it on its micro-browser display.

6.4 MOBILE IP

The systems described so far can provide Internet access as long as the portable device is within a network coverage area. The moment it visits another network, it gets disconnected and the session is terminated. However, people on move need to communicate using only its permanent IP address through Internet even after the change in its current point of attachment to the Internet. Hence the support of mobility in the Internet access is a must. Mobile IP gives a solution allowing people to access Internet on the move without a change in IP addresses.

Mobile IP is an open standard defined by the Internet Engineering Task Force (IETF). It allows users to keep the same IP address, stay connected and maintain ongoing applications while roaming between IP networks. The main objective of this technology is to enable users to keep the same IP address while travelling to a different network (may be operated by a different operator) simultaneously ensuring that a roaming individual could continue communication without sessions or connections being dropped. As it is a network layer solution, it is completely independent of the medium on which it runs. For example, it allows a Laptop to disconnect from a wired Ethernet and switch to a WLAN interface without experiencing a disruption in network service.

Whenever a mobile node moves from one network to the other, its network prefix changes and the node cannot be traced by its fixed IP address. The solution of allocating a new IP address, as soon as the node switches to a different network, is not so efficient. It forces the node to terminate any ongoing communications at the old link and then restart them at the new link. The following section describes a solution for the said problem using mobile-IP.

6.4.1 Architecture

Figure 6.11 shows a typical Mobile-IP architecture. The components of Mobile IP are:

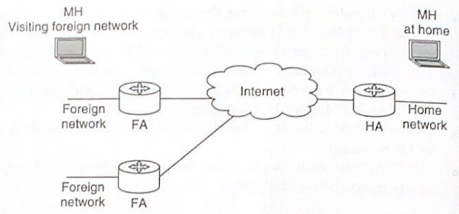


Figure 6.11 Mobile-IP architecture.

Mobile Host (MH): A node that can change its point of attachment to the Internet from one link to another while maintaining any ongoing communications and using only its permanent IP address.

Home Agent (HA): Home Agent is a router with an interface to the MH's home link. It serves as the anchor point for communication with the Mobile Host. Home Agent keeps track of MH's movement maintaining a mobility-binding table described at the end of this section.

Foreign Agent (FA): Foreign Agent is also a router but located in the MH's foreign link. It maintains a visitor list that gives a mapping indicating the location where the data destined for the visiting node (mobile host) can be forwarded.

Both HA and FA periodically advertise their presence through the agent advertisement messages.

Home address: The home address of a mobile host is a permanent IP address assigned to the host. It does not change as the mobile node moves from link to link.

Care-of-address: A care-of-address is the IP address associated with a mobile host visiting a foreign link. It generally changes each time the mobile host moves from one foreign link to another. There are two ways for implementing the concept of care-of-address – the foreign agent care-of-address and collocated care-of-address. The foreign agent care-of-address means the IP address of an FA that has an interface with the foreign link the MH is currently visiting. It can be simultaneously shared by many MHs. On the other hand, the collocated care-of-address is an IP address temporarily allocated to an interface of the MH. When no FA is available on a foreign link, the collocated care-of-address is used. A collocated care-of-address is, therefore, used by only one MH at a time. In a word, the care-of-address is either an address of an FA or an address assigned temporarily to an interface of the MH.

The mobility-binding table in HA contains MH_IP (home address of the MH), c_o_addr (care-of-address of MH) and life_time (a parameter to be discussed later). It maintains an association/mapping between the home/permanent IP address and the care-of-address for routing MH-bound packets. The visitor list in FA contains MH_IP, HA_IP (IP address of HA) and life_time.

6.4.2 How does Mobile IP Work?

The whole process of providing Mobile IP environment is based on three sub-processes, namely Agent discovery, Registration and tunnelling.

Agent discovery

Agent discovery consists of two types of messages—agent advertisement and agent solicitation. To support delivering packets to an MH, the HA and FA periodically broadcast their presence by agent advertisement. All the nodes on the link receive the broadcast. This enables an MH to identify whether it is in home network or it has moved away. If the MH realizes that it is visiting a network other than its home network, it collects care-of-address from agent advertisements.

On the other hand, agent solicitation is sent by an MH which requires agent advertisement instantly and cannot wait for the next periodic agent advertisement. Agent advertisement and agent solicitation are identical to router advertisement and router solicitation of ICMP[3] (Internet Control Message Protocol) Router Discovery Messages. Mobile IP extends ICMP router discovery message to implement agent discovery. In spite of periodic agent advertisements, if a mobile host needs agent information instantly, it can use an ICMP router solicitation message. Any agent receiving this message will then issue an agent advertisement. The event of getting instant agent advertisement is called agent solicitation.

The parameter life_time (lifetime) present in the mobility-binding table and in the visitor list indicates maximum allowable time period between two consecutive advertisements either by an HA or by an FA. Listening the parameter value from an advertisement, an MH realizes how frequently an agent broadcasts its advertisements. In reality during broadcasting, the advertisements may be lost due to the error-prone characteristic of wireless medium. So HA and FA broadcast well before the time period lapses. If an MH is registered with an FA, and fails to listen an agent advertisement within the stipulated lifetime, the MH can realize that either it has moved to a new FA or the desired link is not functioning. In this case, the MH tries to listen advertisement further to take necessary action. If no such advertisement is listened by the MH, it opts for agent solicitation.

[3] ICMP is one of the core protocols of the Internet Protocol suite. ICMP Router Discovery is used by nodes in Internet to detect router running ICMP router discovery.

If for some reasons FAs are busy or no FA is available, the MH itself acts as its own FA by using collocated care-of-address. It can use DHCP[4] (Dynamic Host Configuration Protocol) to contact a service provider in the present network and then obtains the collocated care-of-address. The node sends a request to the DHCP server to lease an address for some period of time.

Registration

The primary purpose of registration for a mobile host is to inform its home agent the current care-of-address collected in the agent discovery phase. Immediately after getting care-of-address, the MH prepares registration request. The registration request includes the MH_IP (IP address of the MH), HA_IP (IP address of its Home Agent) and the care-of-address the MH learns from the FA. The MH sends the registration request to its home agent through foreign agent. The FA checks the validity of the registration request. If the request is valid it forwards the request to its HA. Once the HA knows the care-of-address, it can send packets to the care-of-address to deliver the same to the mobile host.

If the MH detects that it has moved to another network, it sends a new registration request through the new FA. In this case, the mobility-binding table of the HA is updated replacing the MH's old care-of-address. Accordingly the HA forwards the data packets to the new care-of-address.

When the MH returns to its home network, it no more requires mobility status and hence sends a deregistration request to the HA to remove care-of-addresses so far assigned to it.

Tunnelling

As soon as a data packet (datagram[5]) destined to an MH reaches the MH's home network, the HA intercepts the packet. It consults the mobility-binding table to know the care-of-address of the MH. Now the HA constructs a new IP datagram[6] setting the care-of-address as destination IP address. The original IP datagram is put or encapsulated in the payload portion of the newly constructed IP datagram. The use of an outer IP datagram with

[4] The DHCP is an automatic IP address assigning mechanism. This protocol is used by networked nodes to acquire the essential parameters such as IP addresses, subnet masks, etc. to join and get access of an Internet Protocol (IP) network. The protocol basically automates the joining and accessing of IP network by a node by assigning the said parameters.

[5] A datagram is a formatted unit of data carried by a computer network to exchange data between two network nodes.

[6] An IP datagram is used by the network layer to exchange data between two network nodes. It may be of variable size consisting of header and the payload. The header contains addressing and control information whereas the payload contains the data to be sent.

a different destination IP address is known as tunnelling. This outer IP datagram is routed to the FA. The FA in turns decapsulates the same, that is, removes the outer datagram and finds out the MH_IP from the original datagram. The FA then consults the visitor list to find out any such MH_IP in the list. If it is found, the FA delivers the same to the MH. In a word, the HA intercepts IP packets destined to the MH and forwards the same to the MH through the FA. The encapsulation process creates a logical construct called a tunnel between the device that encapsulates and the one that decapsulates. Figure 6.12 shows the encapsulated packet before tunnelling and decapsulated packet at the end of tunnel. The original source and final destination address fields are set to the entry-point and the exit-point of the tunnel respectively.

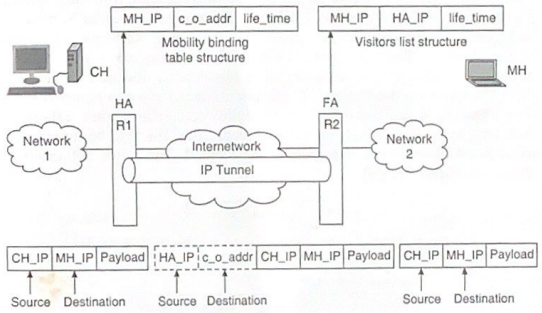


Figure 6.12 Tunnelling.

The tunnelling in Figure 6.12 considers the Correspondent Host (CH) attached to the Network 1 is sending packet {(IP source = CH_IP, IP destination = MH_IP) (IP Payload)} to the MH currently visiting Network 2. The HA intercepts the packet, encapsulates it adding header (IP source = HA_IP, IP Dest = c_o_addr). The c_o_addr is the IP address of FA. Once the packet reaches FA, it decapsulates the packet, i.e. removes the added header and delivers it to the MH.

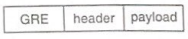
An MH sends packets using its home IP address (HA_IP) effectively maintaining the appearance that it is always in its home network. It forwards packets to the FA that routes those to the correspondent node, the final destination, through HA. In such case, the packets flow through the tunnel established from FA to HA. This is called reverse tunnel.

The default encapsulation process used in mobile-IP is called 'IP encapsulation within IP' or IP-in-IP. In this method an extra IP header is put on the top of the packet to be forwarded. In other words, the original packet (inner packet) to be delivered is encapsulated within the payload

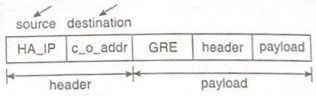
portion of another packet (outer packet). The IP-in-IP encapsulation makes the tunnel appear as a single virtual link to an original packet.

Mobile IP supports two optional encapsulation processes—Minimal Encapsulation and Generic Routing Encapsulation (GRE). Minimum encapsulation also supports only IP network layer protocols. In this encapsulation, some of the redundant fields are discarded from IP-in-IP encapsulation, resulting in less overhead. However, minimal encapsulation works only when an original packet is not fragmented.

In spite of the fact that IP protocols are used as default for mobile IP, many user communities prefer different protocol suites such as Novell Netware or Apple Talk in their organizational network. GRE supports multi-protocol encapsulation. In addition to IP-in-IP encapsulation, the GRE allows encapsulation of packet of one protocol suite into the payload part of a packet of a different protocol suite. For example, consider Network 1 in Figure 6.12 runs on TCP/IP protocol suite where Network 2 runs on Novel Netware. The MH's home network is the Network 1 and visiting network is Network 2. If a packet from Network 2, destined to the MH (MH_IP) arrives at the Network 1, the datagram as shown in Figure 6.13(a) is constructed. It contains a GRE header followed by the original datagram (header, payload) of Novel Netware protocol suite. This will be considered as payload of the outer datagram to be formed. The outer datagram is shown in Figure 6.13(b).



(a) Datagram (Novel Netware protocol suite) preceded by GRE header



(b) Datagram (IP protocol suite) encapsulating datagram in (a)

Figure 6.13 GRE encapsulation.

If tunnel is not set-up perfectly, there is a possibility of formation of a routing-loop[7]. The problem of routing-loop is very serious here as each time the packet reenters the same tunnel, one additional encapsulation occurs causing the packet to grow in size and flow within the network indefinitely. GRE has explicit mechanism for preventing such problem.

[7] The routing loop is a problem arises in network domain. If a link between any two nodes fail and the rest of the nodes in the network are not updated immediately about this failure, the other nodes still try to send packet through the failed link; the node associated with the failed link returns the packet to the sender node intending that the sender would send it via other node and this process will continue making a loop called routing loop.

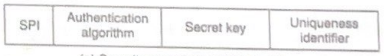
Delivering packets to a mobile host

When a correspondent host (CH) tries to send packets to a mobile host (MH), it uses the home address of MH (MH_IP) as the destination address. The MH examines Agent Advertisement and determines whether it is connected to their home or foreign link. If the MH is still attached to the home network, it receives the packet. Otherwise (the MH has moved to another network), the following steps are performed to delivery the packet to the MH.

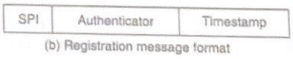
- The HA intercepts packets destined to the MH's home address and consults the mobility-binding table to find out the MH's care-of-address.
- It constructs a new IP packet whose header contains the care-of-address as the destination address and HA_IP as the source address. The payload portion of this newly constructed packet contains the original packet. In other words, the destined packet is encapsulated.
- The HA sends the encapsulated packet.
- The FA in the visiting network, upon receiving the encapsulated packet, removes the header and finds out the MH_IP.
- The FA delivers the packet to the MH after consulting the visitor list.

6.4.3 Security

Mobile-IP provides IP layer security support. The main component of the security support provided by mobile-IP is securing the registration process. During registration a malicious node may pose as FA and forwards a registration request of an MH to an HA. In response of the request, the HA forwards data packet (destined to the MH) to the malicious FA. Mobile-IP provides a way-out of this problem by giving security association between the HA and MH. Security association is a negotiation between two nodes that states how a sender node converts a data prior to transmitting it to the receiving node. This negotiation is made based on the entries in security association table that is known to both the nodes *a priori*. The security association table format is shown in Figure 6.14(a).



(a) Security association table format



(b) Registration message format

Figure 6.14 Implementing secured registration.

An authenticator is computed by using parameter values which an SPI corresponds to. The SPI (Security Parameter Index) is an index that

uniquely identifies a security association between two nodes. The parameters that an SPI corresponds to are the authentication algorithm, secret key and the uniqueness identifier. The commonly used algorithm is keyed MD5 with 128 bits key. The uniqueness identifier is generally implemented by timestamp. In such case each registration message is accompanied by a timestamp. The registration message format is shown in Figure 6.14(b).

The security association that is negotiated between the sender and the receiver is done either manually or by an automated key management protocol. To implement such security association, each registration (request or reply) message should contain authentication extension containing security parameter index followed by an authenticator. The receiver compares the authenticator value it gets from the sender along with registration and one that is computed from the parameters corresponding to the SPI received. The authenticator safeguards the entire registration message. Moreover, if the receiving node realizes the timestamp associated with the registration message is sufficiently reasonable, it accepts the message; otherwise the HA may reject suspicious message thereby protecting replay-attack[8].

The authentication extension between an MH and the HA is mandatory. In addition to this, two other optional authentication extensions exist between the MH and FA and FA and HA. In case of MH and FA extension, when an FA relays a request message to the HA, it removes the extension and send it to the HA. Similarly, when the FA receives response of the requested message, it adds the extension and sends it to the MH.

6.5 CONCLUDING REMARKS

Wireless data service is an area that has witnessed plenty of activities for over a decade. While it remains beyond the scope of this chapter to describe the details of all the advances made, an humble effort has been made to present the general direction in which the evolution unfolds and to mark the milestones thereof.

HSCSD 'uses dedicated circuit-switched' channels and hence it supports a good quality service compared to any packet-switch based data service. Therefore, it is preferable for time-sensitive data transfer such as images, video, etc. The merit of becoming circuit-switched is the demerit in the context of bandwidth usage. GPRS, being a packet-switched solution, is bandwidth efficient compared to HSCSD.

The main competitor of GPRS is CDPD. Both of these provide data service based on packet-switched connection. Data rate and capacity are quite higher in GPRS than CDPD. As CDPD is comparatively older than GPRS, CDPD only supports IPv4 whereas GPRS supports IPv4 and IPv6. GPRS offers both connectionless and connection-oriented service but CDPD offers only connectionless service.

[8] A replay-attack is one type of network attack where a valid data transmission is maliciously intercepted by an adversary and retransmits the data with a delay.

WAP, on the other hand, tried to bring varied Internet contents to cellular handset as well as to other mobile terminals, e.g. PDA (personal digital assistant), palmtop on the top of heterogeneous wireless networks such as GSM, GPRS, CDPD. The goal of WAP is not only to provide data service with higher data rate but also to enable Internet access and other data services with much flexibility in terms of terminals, networks.

The Mobile IP is an extension to the Internet Protocol (IP) proposed by the Internet Engineering Task Force (IETF) that addresses the issue of enabling mobile nodes to stay connected to the Internet regardless of their location and without changing their IP addresses. The other technologies such as GPRS, WAP for providing data service (Internet access) over wireless network can use Mobile IP to enable a mobile node to retain the same IP address and maintain existing communications when away from the home network. Without mobile IP capabilities, the GPRS network would have no way of forwarding packets from the GGSN, destined for the mobile customer, through the SGSN, and ultimate delivery to the GPRS-enabled mobile phone.

BIBLIOGRAPHY

Bates, R. J., *GPRS: General Packet Radio Service*, USA, McGraw-Hill Professional, December 2001.

Bertsekas, D., and R. Gallager, *Data Networks*, 2nd ed., Englewood Cliffs, New Jersey, Prentice-Hall, 1992.

Bevis, D., J. Alvinen (Editor), *The Wireless Application Protocol: Writing Applications for the Mobile Internet*, Addison-Wesley Longman Publishing Co. Inc. USA, 2001.

Budka, K.C., et al., "Cellular Digital Packet Data Networks", Bell Labs Technical Journal, Vol. 2, No. 2, 1997.

"CDPD–Cellular Digital Packet Data, Cell Plan II Specification", prepared by PCSI, San Diego, CA 92121, January 1992.

CDPD Forum, "Cellular Digital Packet Data Specification–Release 1.1", *Technical Report*, Chicago, January 1995.

Cellular Telephone Industry Association, Cellular Digital Packet Data System Specification, Release 1.0, 1993.

Das, S., A. Misra, and P. Agrawal, "TeleMIP: telecommunications-enhanced mobile IP architecture for fast intra domain mobility", *IEEE Personal Communications*, Vol. 7, No. 4, pp. 50–58, 2000.

ETSI (Release'96), High Speed Circuit Switch (HSCSD), stage 2, European Telecommunications Standards Institute, GSM 03.34 v.5.2.0 (1999–2008).

ETSI (Release'98), High Speed Circuit Switch (HSCSD), stage 1, European Telecommunications Standards Institute, GSM 02.34 v7.0.0 (1999–2008).

Evams, H., P. Ashworth, *Getting Started with WAP and WML*, USA, Sybex, April 2001.

Fenton, C.J., et al., *Mobile Data Services*, BT Technical Journal, Vol. 14, No. 3, July, 1996.

General Packet Radio Service (GPRS), Overall description of the GPRS radio interface, ETSI, v 6.0.1, 1998.

General Packet Radio Service (GPRS), Requirements specification of GPRS, ETSI, v 6.0.0, 1998.

Hamalain, J., "High Speed Data service in GSM", *Technical Forum*, Telecom'95, Geneva, October 1995.

Heijden, M., and M. Taylor (Editors), *Understanding WAP: Wireless Applications, Devices, and Services*, USA, Artech House, 2000.

Hoff, S., et al., "A performance Evaluation of Internet Access via the General Packet Radio Service in GSM", *IEEE Vehicular Technology Conference*, Canada, 1998.

Jacobsmeyer, J., "Improving Throughput and Availability of Cellular Digital Packet Data (CDPD)", *Virginia Tech 4th Symposium on Wireless Personal Communications*, USA, pp. 18.1–18.12, 1994.

Kavanagh, A., and J. Beckmeyer, *GPRS Networks*, Osborne Publishing, USA, September 2002.

Mademann, F., "General Packet Radio Service–A packet mode service within the GSM", *International Switching Symposium*, Berlin, 1995.

Mann, S., "The Wireless Application Protocol", *Dr. Dobb's Journal*, 1999.

Pandya, Raj, *Mobile and Personal Communication Systems and Services*, New Delhi, Prentice-Hall of India, 2004.

Perkins, C.E., "Mobile Networking through Mobile IP", *IEEE Internet Computing*, 1998.

Perkins, C.E., S. R. Alpert, B. Woolf, *Mobile IP: Design Principles and Practices*, Addison-Wesley Longman Publishing Co., Inc.USA, 1st ed., 1997.

Saha, D., et al., "Cellular Digital Packet Data Networks", *IEEE Transaction on Vehicular Technology*, Vol. 46, No. 3, August 1997.

Schiller, Jochen H., *Mobile Communications*, New Delhi, Pearson Education, 2007.

Scholefield, C., "Evolving GSM Data Services", *IEEE International Conference on Universal Personal Communications*, San Diego, CA, 1997.

Singhal, S., T. Bridgman, L. Suyranarayan, D. Manuey, J. Chan, D. Bevis, J. Alvinen (Editor), *The Wireless Application Protocol: Writing Applications for the Mobile Internet*, Addison-Wesley Longman Publishing Co. Inc. USA, 2001.

Solomon, J., *Mobile IP: The Internet Unplugged*, Upper Saddle River, New Jersey, Prentice Hall, PTR, 1998.

Stallings, William, *Wireless Communications and Networks*, New Jersey, Pearson Education, 2006.

Zhang, J.Z., and J.W. Mark, "A local anchor scheme for Mobile IP", *Proceedings of International Conference on Performance and QoS of Next Generation Networking*, Nagoya, Japan, pp. 137–156, 2000.

REVIEW QUESTIONS

1. Which system do you consider the first initiative on data service based on GSM network?
2. What are the full forms of HSCSD and CDPD? What is the basic difference between HSCSD and CDPD?
3. How 9.6 kbps data rate in GSM is increased to 14.4 kbps in HSCSD network? How can further increase in speed be achieved in HSCSD?
4. (a) Explain the concept of channel hopping.
   (b) How does the CDPD system exploit channel hopping?
   (c) What is the difference between the emergency hopping and the planned hopping?
   (d) Describe channel assignment strategies in CDPD.
5. What do the modules a mobile end system (M-ES) contain ? Mention the tasks performed by each of the modules.
6. List the responsibilities of Mobile Data Base Station (MDBS). Why is busy/idle status of reverse channel in CDPD important?
7. What are Mobile Home Function (MHF) and Mobile Serving Function (MSF)? Which module in CDPD takes care of mobility management?
8. What are the roles played by the following protocols in CDPD and in which layer of the protocol stack do they function?
   RRMP
   MNRP
   MNLP
9. (a) As compared to GSM, what are the additional network components provided in GPRS to implement packet-switched service?
   (b) Diagrammatically represent GPRS architecture and describe briefly the role of each component shown in the architecture.
10. What is the maximum data rate in GPRS? How can this be achieved?
11. (a) What is P_TMSI?
    (b) Who assigns it to whom?
    (c) When is it assigned?
12. Based on the service domains, what are the three modes of operation a GPRS MS can operate?
13. (a) What are the four categories of traffic supported in GPRS?
    (b) To which categories of traffic do the following applications' traffic belong?
    Web surfing
    File transfer

Downloading
SMS
Video conferencing

(c) Briefly describe the characteristics of each type of traffic.

14. Why is GPRS more attractive than GSM from the subscriber's perspective?
15. (a) What is the need of broadcast channel in GPRS?
    (b) Name the GPRS common control channels.
    (c) Write the role played by each of the common control channels.
16. Describe the tasks of LLC, RLC and MAC layers over air interface in GPRS.
17. Explain how does routing take place for data transfer between a GPRS MS and external data networks.
18. What are the three states of an MS in GPRS ? What is the need of all these states?
19. What is routing area? How do location update and mobility management take place in GPRS?
20. (a) What is PDP context?
    (b) What is QoS negotiation?
    (c) Why does an MS create PDP context?
    (d) How does it take place?
21. Compare GPRS user validation process with GSM.
22. Why are Internet protocols not suitable for Internet access from wireless networks? What may be the probable solutions?
23. What is WAP? What was the objective of WAP forum?
24. List the components of WAP architecture and mention the tasks of each component in brief.
25. (a) Describe the functionality of WAP gateway with the help of a diagram.
    (b) What is WAP PUSH service? How does the gateway participate to enable this service?
    (c) Why does a WAP client maintain agent(s)?
26. (a) What is the difference between WML and HTML?
    (b) Why is this difference important for hand-held devices?
    (c) Why has a scripting language been added to WML?
    (d) How many classes of transaction services are offered by wireless transport protocol (WTP)? Give an example of service of each class.
    (e) Name the layer of the WAP protocol stack at client side that provides security. What levels of security does it provide?
27. Name the three important bearers with which the WAP protocol stack is designed to operate. When is it preferred to use HSCSD as bearer and when is it preferred to use GPRS as bearer?
28. Write the steps of operation for accessing Internet by a WAP terminal/client.
29. What is mobile-IP? Explain the requirements of mobile-IP.

30. List the entities of mobile-IP architecture and briefly describe each of the entities.
31. What are the two different types of destination addresses that can be assigned to a mobile node while it is attached to a foreign network? Under what circumstances would a mobile node choose to use both the types of addresses?
32. (a) What is the mobility-binding table?
    (b) Which entity of mobile-IP does contain this table?
    (c) Name the attributes of the table.
33. Explain how does tunnelling work for mobile-IP using IP-in-IP.
34. What are generic routing encapsulation (GRE) and minimal encapsulations? Explain the requirements of two such encapsulations.
35. (a) What is the need of agent advertisement and agent solicitation? Explain.
    (b) Why registration of a mobile node is essential?
    (c) How does registration take place?
    (d) Write the steps of operation for transferring data from a fixed node to a mobile node and vice versa.
    (e) Where is the need of encapsulation?
36. (a) What level of security does mobile-IP provide?
    (b) What are the probable security threats of mobile node during registration?
    (c) Describe briefly how the threat is taken care of.

7

OVERVIEW OF THIRD GENERATION CELLULAR NETWORK (UMTS)

2G networks were built mainly for voice and slow data transmission. 2.5G such as i-mode (wireless Internet) data services, high-speed circuit-switched data (HSCSD) and GPRS were introduced to provide enhanced functionality. Due to rapid changes in user expectation, they do not meet today's needs. The introduction of 3G cellular systems is a step towards standardization of the next generation mobile cellular networks. The initial aim of 3G systems was to provide a single set of standards that could meet a wide range of wireless applications and provide universal access throughout the world. The distinctions among various devices/services should disappear and a universal personal communicator should provide access to a variety of voice, data and video communication services. The subscriber would be able to get the mobile services from anywhere in the world without replacing his handset or SIM card. However, after prolonged International debate/discussion at political and social level about the ownership (patent/copyright), the target was discarded. Finally, only a 3G standard was taken up.

The 3G standard is accepted to provide fairly high-speed wireless communications that support multimedia, data and video in addition to the voice data. The systems following 3G standard are designed for secure and efficient interconnection with the Internet. Some of the targeted capabilities of 3G, as per the ITU's IMT-2000 (International Mobile Telecommunications for the year 2000) initiative, are—(i) voice quality is to be comparable with the public switched telephone network, (ii) both packet-switched and circuit-switched data services are to be supported, and (iii) various types of mobile terminals are to be accepted.

7.1 3G HISTORY

In many countries, 3G networks do not use the same radio frequencies as 2G. So the mobile operators must build entirely new networks and obtain license for entirely new frequencies. Due to huge investment requirements for additional spectrum licensing fees, introduction of 3G was delayed in some countries. Japan was the first country to introduce 3G (NTT DoCoMo) in 2001 nationally. The transition to 3G in Japan was mostly completed in 2006. In Europe, 3G (UMTS by Telenor) services were introduced during 2002–2003. In the US, the first successful 3G service was launched by Verizon Wireless in 2003. Cellular mobile telecommunication networks are expected to be upgraded to use 3G technologies worldwide by 2010.

7.2 FEATURES OF 3G

The target of 3G cellular system is to offer higher data rate in comparison to 2G system for supporting multimedia data service in addition to voice. It supports both packet-switched (e.g. web browsing), and circuit-switched data service (e.g. real time video). In 3G, voice quality is comparable to Land phone (PSTN). It also supports a wide variety of devices as hand-held terminals.

The 3G system allows transmission of 144 kbps for high-speed mobile systems (e.g. terminal within a motor vehicle) and 384 kbps for slow-speed (e.g. pedestrian with a terminal) mobile systems. For stationary systems (office use), transmission with 2 Mbps rate is supported.

The 3G systems are expected to have greater capacity and better spectrum efficiency. It allows global roaming among different 3G networks. Wide ranges of applications such as person-to-person communications, mobile entertainments, wireless advertising, mobile transactions etc. are supported by the 3G systems.

7.3 RADIO INTERFACES IN 3G

As per the IMT-2000 specification, a set of alternative radio interfaces is recommended for use to provide flexibility for easy transition to third generation from the first and second generation systems. The set of alternative specification are:

- IMT-DS direct spread (W-CDMA)
- IMT-MC multicarrier (CDMA 2000)
- IMT-TC time code (TD-CDMA)
- IMT-SC single carrier (TDD)
- IMT-FT frequency-time (DECT+)

The dominant technology for 3G systems is CDMA. Two out of the above specifications are used by ETSI (European Telecommunications Standard Institute) to develop Europe's 3G wireless standard, namely

Universal Mobile Telecommunications System (UMTS). Both of the specifications are CDMA based—W-CDMA and TD-CDMA. The other CDMA based specification cdma2000 is accepted as a standard in North America and Japan.

7.4 SPECTRUM ALLOCATION FOR 3G

As per ITU recommendation, frequency spectrum allocated for IMT-2000 systems is 1885–2025 MHz and 2110–2200 MHz. However, implementation history of wireless systems varies from sub-continent to sub-continent. In Europe, a part (1880–1900 MHz) of the recommended 3G spectrum is already in use by the DECT system. Out of the rest of the spectrum 1900–2025 MHz and 2110–2200 MHz, UMTS TDD uses 1900–1920 and 2010–2025 MHz whereas UMTS FDD uses 1920–1980 and 2110–2170 MHz. The residual part of the allocated 3G spectrum is used by the satellite services.

CDMA 2000 in Japan uses 1920–1980 MHz and 2110–2170 MHz. In the US, the allocated spectrum for 3G services is already in use by other services. To overcome this conflict, in WRC'2000 (World Radio communication Conference held in 2000), three different bands (800–1000 MHz, 1700–1900 MHz, 2500–2700 MHz) are identified for CDMA 2000 in the US. Unfortunately, again it is observed that some part of the identified band is already in use by other services including Defence. Therefore, reassessing the national need in its entirety, the bands 1710–1770 MHz and 2110–2170 MHz are selected for 3G services.

The following section describes UMTS as an example of 3G systems.

7.5 UMTS

The Universal Mobile Telecommunication System (UMTS) is capable of providing a variety of mobile services to a wide range of global mobile communication standards.

7.5.1 UMTS Services

UMTS supports bearer services, teleservices and supplementary services as in GSM/GPRS. A brief on such services are described below:

Bearer services: In GSM, there are fixed bearer services irrespective of the needs of an application. On the contrary, in UMTS application-specific negotiation on various parameters of the bearer service is allowed. The necessary controlling parameters that characterize the bearer services are traffic type, traffic characteristics, quality of service (QoS), etc. When a connection is being established for an application, the parameter values may be negotiated as per the requirement of the application. Even after the connection is established, if the need arises, the parameter values may further be negotiated. Both connection-oriented and connectionless services are offered.

UMTS supports traffic type for real time and non-real time applications both in circuit switched (CS) and packet switched (PS) domain. Examples of real time applications in CS domain are real time video, audio and speech. File transfer service may be considered as an example of non-real time application in PS domain. To support all the said applications, there are three options to negotiate—constant bit rate, dynamically variable bit rate, and dynamically variable bit rate with a guaranteed minimum. As far as traffic characteristics are concerned, UMTS supports point-to-point and point-to-multipoint communications. The QoS varies depending on delay, bit error ratio, data rate, etc. An application can specify its requirements by requesting a bearer service with the selected parameter values.

In UMTS the offered data rate targets are:

- minimum 144 kbps rural outdoor with a target of 384 kbps
- minimum 384 kbps urban outdoor with a target of 512 kbps
- 2 Mbps indoor and low range outdoor

The traffic types supported (in UMTS) are categorized as:

- conversational or class A (voice, video telephony)
- streaming or class B (multimedia, video on demand)
- interactive or class C (web browsing, database access)
- background or class D (e-mail, SMS, downloading)

Teleservices: It uses the bearer services availed by bottom layers. The example of teleservices are voice/telephony, SMS (short message service) point-to-point, and SMS (cell broadcast) point-to-multipoint Internet access. One additional teleservice, compared to GSM, supported by UMTS is the wideband adaptive multirate codec. With this service, instead of providing services to all classes of data uniformly, different classes of data are treated according to the service needed. For example, conversational data and streaming class of data (class A and B) are more sensitive to the bit errors compared to the background class of data (class D). So the background class of data is provided lower error protection than conversational data and streaming class of data.

Supplementary services: Such services cannot be offered as separate service. These are offered in association with the basic telecommunication services. One supplementary service may be associated with many telecommunication services and one telecommunication service may have many associated supplementary services. The example of supplementary services are call forwarding, call waiting, call hold, call barring, etc. In comparison to GSM, UMTS supports an additional supplementary service called multicall. This multicall service enables a subscriber to connect and maintain two simultaneous calls in CS domain. For example, if a UMTS terminal UE (user equipment) has access to the multicall service, it can set up a video call in addition to a voice call.

From the discussion on various data services provided by UMTS, it is seen that in addition to enhancing data rate, the new features to UMTS

are multimedia data service, point-to-multipoint communication, wideband adaptive multirate codec, and multicall facility. Moreover, it supports creation of a virtual home environment (VHE) for visiting subscribers. The introduction of VHE allows the subscribers to get desired customized features consistently irrespective of the location of the subscribers, provided the subscribers have compatible terminal and belong to compatible network. Another useful addition is open service access (OSA). In GSM, the applications for available services have been developed and maintained by the network operators themselves. On the contrary, OSA allows applications developed by the third party vendors.

7.5.2 UMTS Architecture

Figure 7.1 shows a UMTS network consisting of three interacting domains—Core Network (CN), UMTS Terrestrial Radio Access Network (UTRAN) and User Equipment (UE). Although the basic Core Network architecture for UMTS is similar to that of GSM network with GPRS, the UMTS uses entirely new radio interface. All the equipments used in GSM/GPRS radio interface therefore have to be modified for UMTS operation and services. The new radio network in UMTS is UTRAN. It provides the interface access method for User Equipment.

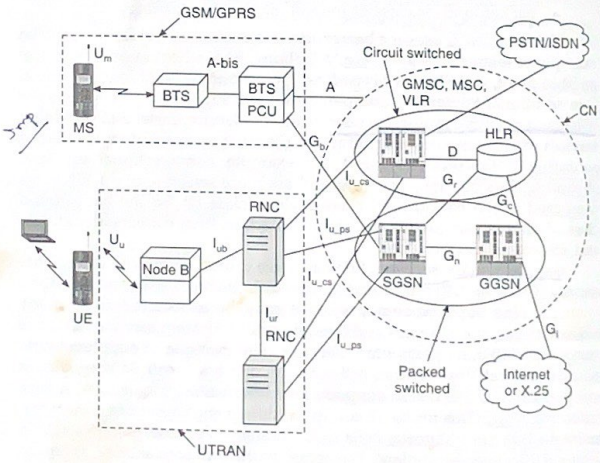


Figure 7.1 UMTS architecture.

Core network

The Core Network (CN) is divided into circuit-switched and packet-switched domains. Circuit-switched elements are the Mobile Switching Centre (MSC), Visitor Location Register (VLR) and Gateway MSC (GMSC). On the other hand, the packet-switched elements are Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN). The network elements EIR (Equipment Identity Register), HLR (Home Location Register), VLR (Visiting Location Register) and AUC (Authentication Centre) are shared by both the domains. The main functions of the core network are to provide switching and routing for user traffic. It also contains the databases and network management functions, functions responsible for intersystem handover and location management. The CN communicates with the UTRAN via I_u interface (Figure 7.1). This is comparable to GSM's 'A' interface. The MSC in CN communicates with RNCs via $I_{u\text{-}CS}$ interface whereas the SGSN communicates with an RNC via $I_{u\text{-}PS}$ interface.

The functionalities of MSC as well as SGSN in GSM/GPRS are changed when these are used to UMTS. In a GSM system, the MSC handles all the circuit-switched operations such as connecting two subscribers through the network. The SGSN/GGSN handles all the packet-switched operations and acts as the gateway to other fixed and wireless data network while transferring data between the UMTS and other data network.

UTRAN

The UTRAN comprises of Node-Bs and RNCs. Node-B in UTRAN corresponds to GSM base station whereas the control equipment equivalent to a Base Station Controller (BSC) for Node-B's is called Radio Network Controller (RNC). A group of Node-Bs is connected to an RNC. The RNC with which a Node-B is connected is called controlling RNC (CRNC) of that Node-B. The set of the Node-Bs along with their CRNC is referred to as the Radio Network Subsystem (RNS). The UTRAN consists of several such RNSs.

The UTRAN is connected to the CN through I_u interface that is similar to 'A' interface in GSM. To be more specific, each RNC in UTRAN connects the MSC of CN by $I_{u\text{-}CS}$ interface and connects the SGSN of CN by $I_{u\text{-}PS}$ interface. It signifies that the packet-switched data is transmitted through $I_{u\text{-}PS}$ interface and circuit-switched data is through $I_{u\text{-}CS}$ interface. An RNC controls more than one Node-Bs over I_{ub} interface and two RNCs are connected through I_{ur} interface. Although RNCs are equivalent to BSCs in GSM, I_{ur} interface has no peer in GSM. This interface facilitates handling of entire radio resource management and eliminates the burden from CN. The UE is connected to Node-B over W-CDMA U_u interface.

Node-B is the physical unit of radio transmission/reception to/from UEs. It can be collocated with GSM BTS to reduce implementation costs. A Node-B connects one or more antennas that corresponding to one or more cells. Moreover it measures power level of the radio links to the UEs attached over air interface and sends the parameter values to its CRNC.

Based on these information, the CRNC adjusts the power of the radio signal at Node-B and UE. The Node-B also measures signal strength of the connection with UE and computes frame error rate (FER). The measured values are then sent to the CRNC, enabling it to take handover decision. The major functions of Node-B are summarized below:

- Air-interface Transmission/Reception
- Modulation/Demodulation
- Measurement of power of the radio links to the UEs
- Close loop power control
- Measurement of signal strengths of the radio links to the UEs

The RNCs perform radio resource management. Wide band CDMA (WCDMA) technology, a new interface compared to GSM, is used as UTRAN air interface. WCDMA has two basic modes of operation—Frequency Division Duplex (FDD) and Time Division Duplex (TDD). The RNCs assign unique WCDMA code for each UE enabling the RNC to segregate the data sent by a UE from the data received from all the UEs around the Node-B under the control of the RNC. To run a CDMA system, it is essential to keep the interference below a threshold. The RNC computes the traffic in each cell and accordingly decides to accept/reject a new call. RNCs also assist in soft handover of the UEs between cells, RNCs and RNSs. The important functions of RNC are listed below:

- Radio Resource Management
- Admission Control
- CDMA Code Assignment
- Power Control
- Handover Control

User equipment

UE (User Equipment) is a wireless device such as cell phone, PDA (Personal Digital Assistant), etc. As mentioned earlier, the UE is connected to a Node-B via the WCDMA radio interface U_u. At any point of time, a UE may be attached to more than one cell/Node-B. If there is a need that the 3G handset/UE is to be backward compatible, that is, capable to access 2G networks too, technical complexity increases. With the increased complexity in design, the size, weight and cost also increase. In UMTS, the UEs can roam into the GSM/GPRS network and therefore incur the additional complexity. In Europe, the manufacturers and network operators wanted multi-mode 3G phones, which would operate on both the 3G and 2G networks. This multi-mode operation added to the complexity in designing the handset. As a result, early European WCDMA sets were comparatively larger and heavier than Japanese WCDMA sets.

Terminals have many different types of identities. Most of these UMTS identity types are taken directly from GSM specifications. In the CS domain, a UE is identified by IMSI (International Mobile Subscriber Identity) and TMSI (Temporary Mobile Subscriber Identity) as an MS is identified in GSM. Similarly in the PS domain, a UE is identified by IMSI and P_TMSI (packet TMSI) as in GPRS.

The UMTS uses USIM (User Service Identity Module) that has same physical characteristics as the GSM SIM card. All the essential data for authentication and access to the UMTS network are stored in the USIM card. To keep personal data, a GSM SIM card has storage capacity in the order of kilobyte whereas the USIM storage capacity in the order of megabyte. This extended storage capacity is for accommodating one or more user profile and downloading of new applications.

As a UE has to communicate different network elements such as Node-B, RNC and CN, it participates in performing several tasks done by the corresponding network elements. For example, as part of UE and Node-B communication, the UE helps Node-B in signal strength measurement and power control measurement. As part of UE and RNC communication, the UE cooperates in handover process and radio resource management. Further, a UE seeks services from the network from time to time and takes part in bearer negotiation as part of UE and CN communication.

Based on the service domains, a UE can operate in one of the following three modes:

- PS and CS modes of operation: The UE is attached to both the PS and CS domains and is capable of simultaneously operating PS services and CS services.
- PS mode of operation: The UE supports only PS connection.
- CS mode of operation: The UE is attached to the CS domain only.

7.5.3 Mobility Management

Mobility management is the task that enables a subscriber of a wireless network to communicate (voice or data) on move. Location tracking is an essential task of mobility management by means of which the network keeps track of a mobile station to establish connection for voice or data communication to and from the mobile station. Handover is the other major task of mobility management for transferring an ongoing communication session from one controlling element of the network to another. These two tasks are described in this section.

Location tracking

In UMTS, the network maintains location information of each UE. The information can be updated by a UE whenever there is a change of location. The change of location is determined whenever the UE starts receiving a broadcast message from the network that is different from the message received previously.

To implement such location tracking mechanism, UMTS introduces the concept of grouping of cells defined as URA (UTRAN routing area), RA (routing area) and LA (location area) based on coverage by Node-Bs,

RNC and SGSN respectively. A group of cells covering the area of one or more Node-Bs under one RNC is called URA. One RNC may have several URAs. One or more URAs under one RNC may form one RA. One RNC may have one or more RAs. Similarly one or more RAs managed by the same SGSN are called LA (location area).

In CS domain, a UE is tracked by the CN (MSC/VLR) to find out in which LA the UE is located. In PS domain, location tracking is done by both the CN (SGSN) and UTRAN. The SGSN tracks a UE at RA level, that is, under which serving RNC it is located whereas UTRAN tracks a UE at cell level. As far as tracking by CN (SGSN) is concerned, the UE has two states—idle (CN) and connected. On the other hand, for tracking by UTRAN, the UE has three states, namely idle (UTRAN), cell-connected and URA-connected.

When a UE originates a data access, a PS signalling connection is established between the UE and SGSN and the UE enters to the connected state. This signalling also triggers the connection establishment between the UE and its serving RNC (UTRAN) and the UE enters to cell-connected mode. The serving RNC tracks the cell in which the UE is located and the data transfer takes place between the originating UE and the external networks. So data communication is only possible in connected and cell-connected mode. In this state, if data communication does not take place for a predefined time interval, the UE is changed to URA-connected. The UE still remains connected but it is tracked to know its URA. However, if there is a need for transferring data to and from the UE, it goes back to the cell-connected mode. As soon as the data transfer is finished or there is no data communication for a predefined period, the UE again enters into idle (CN)/idle (UTRAN) state. A UE in URA_connected mode first switches to cell_connected and then to idle (UTRAN) mode.

Handover

The UMTS mainly supports soft handover ('handoff strategies', Section 2.7.1). The type of soft handover is determined based on the movement of mobile stations (UE) between coverage areas.

Soft handover

In soft handover, at least one radio link exists between a UE and the UTRAN at any point of time. There can be three types of soft handover—inter Node-B and intra RNC, inter RNC and intra SGSN and inter SGSN.

If a UE moves from a cell under one Node-B (NB_{old}) to a cell under another Node-B (NB_{new}), where both the NB_{old} and NB_{new} are under the control of same RNC, a soft handover is performed. It is an inter Node-B and intra RNC type of handover. Based on the signal quality received from the UE, whenever its RNC decides a link must be established between the UE and a new Node-B, the RNC instructs the NB_{new} for setting up the link. The NB_{new} then starts receiving data from the UE. During this handover, the RNC receives the signal from both the NB_{old} and NB_{new} and verifying the quality of the signals received from the two, selects one with better quality. It then sends the packet to its CN. In future, if the UE moves away from one (say NB_{old}) of these Node-Bs, identified based on the measured data from the UE, the controlling RNC decides to detach the radio links between the UE and NB_{old}. Accordingly the RNC instructs the UE to deactivate the link and the NB_{old} releases the radio resources getting an intimation from the RNC.

The inter RNC and intra SGSN soft handover is performed when a UE moves from the coverage of one Node-B (NB_{old}) to another Node-B (NB_{new}), where these two Node-Bs are controlled by the two different RNCs belonging to the same SGSN. Based on the data received from a UE, whenever the controlling RNC (RNC_{old} or serving RNC) decides that a radio link needs to be established between the UE and a new RNC (RNC_{new}). The RNC_{old} requests the RNC_{new} through I_{ur} interface to provide necessary radio resources. If resources are available, the RNC_{new} instructs the Node-Bs under its control to initiate receiving data from the UE. During this handover, the RNC_{new} sends the data received from UE to the RNC_{old} and the RNC_{old} in turn accepts the better quality packets out of the received packets from the two Node-Bs under RNC_{old} and RNC_{new}. Finally, the RNC_{old} sends the accepted packets to CN. When the UE moves away from one (say RNC_{old}) of the RNCs, the radio link between UE and the RNC_{old} is removed and the RNC_{new} is set as the RNC_{old} or serving RNC. This event is called serving RNC relocation.

The inter SGSN soft handover is performed due to the movement of a UE from the coverage of one RNC to another RNC under the control of different SGSN. Based on the data received from a UE, whenever the controlling RNC (RNC_{old} or serving RNC) decides that a radio link needs to be established between the UE and another RNC (RNC_{new}) that belongs to a different SGSN ($SGSN_{new}$), the RNC_{old} reports it to its controlling SGSN ($SGSN_{old}$). The $SGSN_{old}$ contacts the $SGSN_{new}$ through G_n interface and sends necessary information related to the UE. The $SGSN_{new}$ instructs the RNC_{new} to be ready to accept data from the UE and informs the same to the $SGSN_{old}$. The $SGSN_{old}$ then instructs the RNC_{old} to forward the buffered data packets (if any) received from the UE, to the RNC_{new}. Finally, the radio link between the UE and the $SGSN_{old}$ is snapped and the $SGSN_{new}$ and RNC_{new} take charge of the $SGSN_{old}$ and RNC_{old} respectively.

Softer handover

Softer handover is a special case of soft handover. It occurs when a UE moves from one cell to another cell both under the control of same Node-B. It is an intra Node-B handover. Based on the data received from the UE, whenever an RNC decides that a different link must be established between the UE and the Node-B, it instructs the Node-B for setting up the link. The Node-B then starts receiving data from the UE on its new antenna set for covering the cell the UE currently moved to.

Hard handover

Over and above soft handover, the UMTS also handles hard handover. Hard handover means that all the old radio links to a UE are removed before the new radio links are established. This is done to support the need of change of the carrier frequency (inter-frequency handover). The handover in GSM (TDMA) system described in Section 3.3.4 is the example of hard handover. In UMTS (CDMA), normally for the implementation of handover to other system such as GSM, a hard handover addressing technique is to be adopted.

UMTS supports global roaming. In most of the countries, UMTS mobile stations (UE) support dual mode (UMTS/GSM) operation. The UEs are backward compatible, that is, a UE can be used in GSM network. During a call if a UE moves from the UMTS network to GSM network, a hard handover occurs. However, this handover is transparent to the user.

7.5.4 Protocol Stack

Figure 7.2 illustrates the UMTS signalling protocol stack both for CS and PS domains. The bottom three layers are the PHY (physical) in layer 1, MAC (Medium Access Control) and RLC (Radio Link Control) in layer 2, and RRC (Radio Resource Control) in layer 3. The physical layer protocol works over the air interface between UE and Node-B. The rest of the bottom layer protocols MAC, RLC and RRC work between UE and RNC.

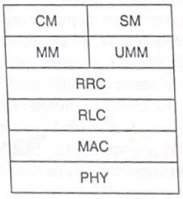


Figure 7.2 UMTS signalling protocol stack.

In CS domain, the top two layers are called CM (call management) and MM (mobility management). The CM and MM protocols are defined between UE and MSC. The CM is responsible for management of call establishment and call release. This also takes care of the supplementary services such as call forwarding, call barring, etc. On the other hand, the MM takes care of the tasks of CS-domain mobility management of UE. The example task in CS-domain mobility management is LA (location area) update.

In PS domain, the top two layers are responsible for SM (session management) and UMM (UMTS mobility management). The SM and UMM are defined between UE and SGSN.

In CS domain, voice communication among UE, Node-B, RNC and MSC is established as in GSM/GPRS. The steps of GSM call establishment can be found in Section 3.3.6. Hence in this section, signalling protocol stack used in PS domain is only elaborated.

Figure 7.3 shows the protocol stack for communication between UE and UTRAN as well as UTRAN and CN. The UMM/SM are the top-level protocols. As in the MM of CS-domain, the UMM manages the tasks of mobility functions such as RA (routing area) update. The RA update is described in Section 7.5.3.

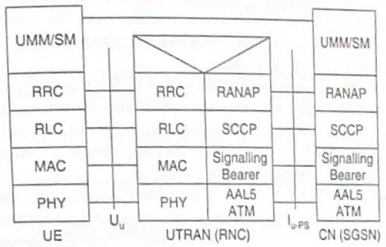


Figure 7.3 UMTS protocol stack (PS domain).

The entity that manages the establishment, maintenance and release of packet connection or session is called session management (SM). The states of session management are either active or inactive. Packet data communication is possible only when SM is in active state. The protocol followed for such a data transfer is called PDP (Packet Data Protocol). The PDP supports data transfer from different packet data networks such as Internet (IP), X.25, etc. It uses various parameters necessary to describe a packet data connection, e.g. allocated IP addresses of both the source and destination, connection type, QoS class, etc. The SM uses these PDP parameters as the PDP context which is defined in the UE, SGSN and GGSN. It can be activated, deactivated or modified. As the PDP is activated/deactivated, the SM is also activated/deactivated. A PDP context can be modified if both the SM and PDP contexts are in active states and is established for one packet-switched service with selected QoS parameter values. Therefore, two different service classes such as class B (e.g. video-on-demand) and class C (e.g. Internet browsing) require two different PDP contexts.

Whenever data transfer takes place between a UE and the CN (core network), a PDP context needs to be created/activated in UE, GGSN and SGSN. It enables exchanging IP packets between UE and the network. To establish an end-to-end connection for such data transfer, the PDP context consists of several parameters such as PDP type (e.g. IPv4), PDP address type (e.g. IP address), QoS profile request (e.g. maximum bit rate 128 kbps

for Internet access), QoS profile negotiated, etc. For example, if the PDP type is IP, PDP context allocates an IP address to the UE.

The RRC (Radio Resource Control) protocol ensures reliable connection between a UE and the UTRAN. It also manages radio resource between UE and UTRAN. The management of radio resource includes establishment, maintenance and release of radio bearers. The other responsibilities of RRC are broadcast of system information from the network to all UEs, paging, power control, encryption, Integrity protection, etc.

An RLC (Radio Link Control) and MAC (Medium Access Control) constitute layer 2 protocol at U_u interface. The RLC performs segmentation and reassembly, padding, and error correction. On the other hand, the major functions of MAC are mapping of channels, measuring traffic volume, handling priority of data flow, etc.

The PHY (physical) layer defines access to the transmission medium. This layer is responsible for error detection and reporting to its upper layer. It takes care of CRC (Cyclic Redundancy Check), coding/decoding, modulation/demodulation, interleaving/de-interleaving. Physical layer also implements functions related to soft handover.

Two protocols SCCP (Signal Connection Control Part) and RANAP (Radio Access Network Application Part) work at $I_{u\text{-}PS}$ interface. These manage the connection between UTRAN and CN. SCCP ensures reliable connection between the UTRAN and SGSN.

RANAP takes care of establishment, modification and release of RABs (Radio Access Bearer). It enables mobility management signalling transfer between the UE and the CN (SGSN). When a UTRAN informs the CN about the position of a UE under the control of a RNC of that UTRAN, the CN manages this mobility management data. This data is not interpreted by the UTRAN. The CN uses the protocol to trigger the UTRAN for broadcasting paging message to track a UE. The RANAP is also responsible for serving RNC relocation as described in Section 7.5.3.

SCCP is the protocol of SS7 (Section 3.3.1) at layer peer to the network layer of OSI. It provides both connectionless and connection-oriented services. At $I_{u\text{-}PS}$ interface, on connection-oriented link, it separates each UE and establishes a connection-oriented link for each of them.

7.5.5 Establishment of a CS Communication

In CS domain, there are three possible connections (as in GSM) involving a UMTS network—Land to Mobile (L_M), Mobile to Land (M_L) and Mobile to Mobile (M_M).

UMTS-call originated by a land phone

In this case, the call is originated by a land phone and destined to a UE that is L_M connection is to be considered. When a call originates from land phone, it is forwarded to the MSC through PSTN. The MSC then checks (from HLR/VLR) whether the called UE is available in the network.

If it is available, a traffic channel is assigned to the UE and the call is set-up after exchange of a number of related messages among MSC, RNC and UE. The activities of the network elements to track the UE and set up the call are sum up in the following steps:

- RNC
  - Periodically broadcast system information
- MSC/VLR
  - Request RNC to track the called UE
- RNC
  - Broadcast on paging channel to track the UE
- UE
  - Detect the page
  - Reply the page and requests for connection (RRC connection request) to the network
- RNC
  - Grant the connection request by assigning a channel to the UE
- UE
  - Send RRC connection set-up completion information
- RNC
  - Relay the connection set-up completion to MSC/VLR
- MSC/VLR
  - Send request to RNC seeking authentication of UE and data for securing communication
- RNC
  - Forward the requests to UE
- UE
  - Send data for authentication and for securing communication
- RNC
  - Forward the response data to MSC/VLR
- MSC/VLR
  - Send connection set up request
- RNC
  - Forward the connection set-up request to the UE
- UE
  - Confirm the connection set-up
- RNC
  - Relay the confirmation to MSC/VLR
- MSC/VLR
  - Instruct RNC to assign radio bearer (necessary channels for voice communication)
- RNC
  - Assign channel to UE
- UE
  - Send channel assignment response
  - Generate alert message (ring tone)

UMTS-call originated by a UE

The UE originated call may be destined to either a land phone or to another UE that is M_L and M_M connections are to be set up. When the call originates, it acquires a traffic channel by exchanging messages among MSC, RNC and UE. The following steps summarize the activities of the originating UE, RNC and MSC while acquiring a traffic channel for the UE to set up a call.

- RNC
  - Periodically broadcast system information
- UE
  - Send connection requests (RRC connection request) to the network
- RNC
  - Grant the connection request by assigning a channel to the UE
- UE
  - Send RRC connection set-up completion information
  - Send service request
- RNC
  - Relay service request to MSC/VLR
- MSC/VLR
  - Send request to RNC seeking authentication of UE and data for securing communication
- RNC
  - Forward the request to UE
- UE
  - Send data for authentication and for securing communication
- RNC
  - Forward the authentication data to MSC/VLR
- UE
  - Send connection set-up request
- RNC
  - Relay the connection request to MSC/VLR
- MSC/VLR
  - Grant the request
  - Instruct RNC to assign radio bearer (necessary channels for voice communication) to the UE
- RNC
  - Grant the radio bearer request to the UE
  - Assign channel to UE
- UE
  - Send channel assignment response

For the M_M connection, the originating UE gets permission of network access and acquires traffic channel by exchanging messages among MSC, RNC and UE as shown in 'UMTS-call originated by a UE'. On the other hand, the called UE gets permission to access the network channel only after the called UE is tracked. The called UE acquires channel after message exchanges among MSC, RNC and UE as the steps shown in 'UMTS-call originated by a land phone'.

7.5.6 Establishment of a PS Communication

In PS domain, data transfer takes place between a UE and the external data network via the CN. In order to establish data transfer, a process is originated either by the UE or by the CN.

PS data transfer originated by UE

If a UE intends to establish a secure connection to the SGSN for sending data (e-mail), accessing external network (Web browsing), etc., the UE originates the process of transfer. The activities of the UE and the other network elements are described below.

- RNC
  - Periodically broadcast system information
- UE
  - Send connection requests for connection (RRC connection request) to the network
- RNC
  - Grant the request by assigning a channel to the UE
- UE
  - Send RRC connection set-up completion information
  - Send service request
- RNC
  - Relay service request to SGSN
- SGSN
  - Send request to RNC seeking authentication of UE and data for securing the communication
- RNC
  - Forward the authentication requests to the UE
- UE
  - Send relevant data for authentication and for securing communication
- RNC
  - Forward the UE's response of authentication request to SGSN
- UE
  - Send PDP context activation request
- RNC
  - Relay the activation request to SGSN
- SGSN
  - Grant the request
  - Instruct RNC to assign radio bearer (channels necessary for PS communication) to the UE

- RNC
  - Grant the radio bearer request to the UE
  - Assign channel to the UE
- UE
  - Send channel assignment response
  - Start uplink data transfer

PS data transfer originated by CN

If a SGSN receives a downlink packet for a UE, it sends a paging request to the UTRAN/RNC. The RNC in turn activates the UE to start the process for data transfer between the UE and the CN. A brief account of the actions taken by different elements is given below.

- SGSN
  - Receive a downlink packet
  - Send a paging message to UTRAN/RNC to track the UE for which the SGSN received the packet
- RNC
  - Broadcast on paging channel to forward the paging request
- UE
  - Detect the page
  - Reply the page and requests for connection (RRC connection request) to the network
- RNC
  - Grant the connection request by assigning a channel to the UE
- UE
  - Send RRC connection set-up completion information
  - Send paging response
  - Send service request
- RNC
  - Relay the service request to SGSN
- SGSN
  - Send request to RNC seeking authentication of UE and data for securing communication
- RNC
  - Forward the requests to UE
- UE
  - Send data for authentication and for securing communication
- RNC
  - Forward the authentication data to SGSN
- SGSN
  - Send PDP context activation request
- RNC
  - Forward the activation request to the UE

- UE
  - Activate the PDP context
  - Confirm the PDP activation completion
- RNC
  - Relay the confirmation to SGSN
- SGSN
  - Instruct the RNC to assign radio bearer (necessary channels for PS communication)
- RNC
  - Assign channel to the UE
- UE
  - Send channel assignment response
- RNC
  - Forward channel assignment confirmation to the SGSN
- SGSN
  - Start downlink data transfer

7.5.7 Power Control

In a CDMA system, it is important that the base station (BS) receives all the UEs at around the same power level. But due to near–far effect (Section 2.5) UEs that are located farther from Node-B, will be received with lower power level. To overcome this problem, both the UEs and the serving Node-B need to adjust their power level efficiently. The power level of the UE is adjusted by two techniques—open loop and closed loop. Open loop technique is to estimate the transmitted power requirement of the UEs. The estimation is done by the UEs measuring the received signal strength from the Node-B. It is employed during set up phase, that is, before communication starts between the UE and Node-B. On the other hand, when the communication is established fully, closed loop technique is used. In each time slot, signal strengths received from UEs at Node-B are measured and a request is sent, if necessary, to the corresponding UE to adjust power level. The Node-B instructs the nearby UEs to reduce their transmitted power, and to increase those who are farther away. In this way the received signal from all UEs will maintain an approximately the same power level.

As the signals transmitted by different Node-Bs are not orthogonal to each other, it is probable that the UEs under the control of one Node-B can also receive signals from their neighbouring Node-Bs. Therefore, the power level of a Node-B are also to be kept as minimum as required by the UEs under its control.

7.5.8 User Validation in UMTS

In UMTS, unlike GSM, not only the network authenticates the UE (MS in GSM) but also the UE authenticates the network. Authentication is

performed for each location update and call-origination (CS domain)/ attempts-for-data-communication (PS domain) by UE. When a UE either requests for location update or desires to originate voice or data communication, the MSC or SGSN sends authentication request along with IMSI of the UE to the HLR/AUC for authenticating the UE. In response to the authentication request, the HLR/AUC searches for the record related to the UE identified by the IMSI. Accordingly, the HLR/AUC generates 5-tuple authentication vector (AV) consisting of RAND (a random number), AUTN (authentication token), XRES (expected response), CP (ciphering key) and IK (integrity key). Now, the CN (MSC in CS domain or SGSN in PS domain) sends RAND and AUTN to the UE. If the AUTN is acceptable to the UE, it authenticates the network. If the UE authenticates the network, it generates a RES (response) and sends it to the CN. After comparing RES with XRES, if the CN finds the match, the UE authentication is completed.

Further, the UE computes CK and IK with the received AUTN. When a UE sends data, the CK and IK are used for ciphering and computing integrity functions respectively. Similarly, the CN sends CK, IK received from HLR/AUC during delivery of AV to the UTRAN for data ciphering and integrity.

BIBLIOGRAPHY

Bic, J.C. and E. Bonek, *Advances in UMTS Technology*, London, Kogan Page Science, 2002.

Carsello, R.D., et al., "IMT-2000 Standards: Radio Aspects", *Personal Communications Magazine*, Vol. 4, No. 4, August 1997.

Chaudhury, P., W. Mohr, and S. Onoe, "The 3GPP proposal for IMT-2000", *IEEE Communications Magazine*, Vol. 37, No. 12, pp. 72–81, 1999.

Dasilva, J.S., et al., "European Third-Generation Mobile Systems", *Communications Magazine*, Vol. 34, No. 10, October, 1996.

ITU-T Recommendation Q.1711, "Network Functional Model for IMT-2000", *International Telecommunication Union*, Geneva, 1999.

Lescuyer, P., and F. Bott, *UMTS: Origins, Architecture and the Standard*, London, Springer, 2004.

Muratore, F., *UMTS: Mobile Communications for the Future*, West Sussex, England, John Wiley, 2001.

Nielen, Van, "UMTS: A Third Generation Mobile System", *IEEE Third International Symposium on Personal, Indoor & Mobile Radio Communications*, Boston, pp. 17–21, 1992.

Prasad, R., W. Mohr and W. Konhauser (Editors), *Third Generation Mobile Communication Systems*, Artech House, 2000.

Smith, Clint, *3G Wireless Networks*, New York, McGraw Hill/Osborne, 2006.

Tabbane, S., "Location management methods for third-generation mobile systems", *IEEE Communications Magazine*, Vol. 35, No. 8, pp. 72–84, 1997.

Universal Mobile Telecommunications Forum, "IMT-2000 Spectrum Requirements", *Special Ad Hoc Group Report*, UMTS Forum, London, 1997.

Walke, B., P. Seidenberg and M. P. Althoff, *UMTS: The Fundamentals*, West Susses, England, John Wiley, 2003.

Zeng, M., A. Annamalai and V. Bhargava, "Harmonization of Global Third-Generation Mobile Systems", *IEEE Communications Magazine*, 2000.

REVIEW QUESTIONS

1. What are the key features that distinguish 3G cellular systems from 2G cellular systems?
2. What is the main reason for delay in launching 3G in some countries?
3. Name the alternative radio interfaces recommended for 3G system by IMT-2000.
4. Name the 3G system introduced in Europe.
5. What are the different data rates supported in UMTS for users with different speeds?
6. How many types of services are supported in UMTS? As compared to GSM, mention the additional services (if any) provided in UMTS in each type of service.
7. In which class of UMTS traffic type the following applications belong:
   (a) web browsing
   (b) voice
   (c) e-mail
   (d) video-on-demand
8. What are the parameters that characterize the bearer service?
9. Discuss the concept of application-specific QoS negotiation.
10. What is the full name of UTRAN? What are the components of UTRAN? Describe the responsibilities of each component of UTRAN.
11. What components of core network (CN) in UMTS perform PS domain and CS domain functionalities?
12. What is USIM? Compare USIM and SIM (GSM).
13. What are the different modes of operation based on service domain a UMTS terminal can operate?
14. What do you understand by hard handover? When does hard handover take place in UMTS? In which way is it different from soft handover?
15. How many types of soft handover does UMTS allow?

16. What is serving RNC relocation? In which soft handover this RNC relocation occurs?
17. In which service domain session management (SM) and UMTS mobility management (UMM) work?
18. What is PDP context? What role does it play in order to transfer data between a UMTS terminal and external data network?
19. Write message exchange that takes place among different components of the UMTS network when a UMTS terminal initiates data transfer to the external data network.
20. Why are both open loop and closed loop power control essential in UMTS?
21. In which way UMTS validation is different from validation in GSM? Explain.