CHAPTER 9

MOBILE DATA NETWORKS

9.1 Introduction

9.1.1 What Is Mobile Data?
9.1.2 Independent Mobile Data
9.1.3 Shared Mobile Data
9.1.4 Overlay Mobile Data

9.2 The Data-Oriented CDPD Network

9.2.1 What Is CDPD?
9.2.2 Reference Architecture in CDPD
9.2.3 Mobility Support in CDPD
9.2.4 Protocol Layers in CDPD

9.3 GPRS and Higher Data Rates

9.3.1 What is GPRS?
9.3.2 Reference Architecture in GPRS
9.3.3 Mobility Support in GPRS
9.3.4 Protocol Layers in GPRS

9.4 Short Messaging Service in GSM

9.4.1 What is SMS?
9.4.2 Overview of SMS Operation

9.5 Mobile Application Protocols

9.5.1 Wireless Application Protocol (WAP)
9.5.2 i-Mode

Questions

Problems

9.1 INTRODUCTION

In Part III of the book we provide detailed examples of important wide area wireless systems. In the previous two chapters, we provided examples of TDMA and CDMA voice-oriented networks. This chapter provides an overview of the lower speed wide area wireless data services that we refer to as *mobile data networks* or *wireless wide area networks.* In the remainder of the book, our emphasis is on the data-oriented networks and Part IV, following this chapter, is devoted to the details of broadband and ad hoc wireless local networks. Wireless data networks are becoming increasingly important in the light of wireless access to the Internet, development of wireless-oriented consumer products, and wireless home networking. By far, voice services have been the major revenue generators for a long time, both in the wireless and wired networks. Long distance and local telephony have provided the biggest revenue for telecommunications networks over the last century. In the 1990s, the emergence of the PC and the Internet have shifted the focus toward data services over the same media as voice services and alternative media like coaxial cable and wireless as well. The driving applications and revenue generators have included content-based services that draw revenue from advertisements, electronic commerce (E-commerce), and, more recently, mobile commerce (M-commerce). Although they were started in 1983, until recently mobile data services were considered expensive, unreliable, and slow. In the past few years, despite all market predictions of failures since the original inception of this industry, the same services that have drawn attention in the wired data world have drawn a new surge of extensive attention in wireless data services. In the past year or so the unpredicted success of SMS in Finland and Japan has catalyzed this new growth of the data oriented industry. Cellular service providers, who once were only focused on voice applications as the main source of income, today envision the future as being comprised of more and more data applications. The mobile data industry that started with ARDIS (in 1983) as a private network for IBM [PAH94] is now being assimilated as a public service into the next generation of the cellular networks and plays an important role in shaping the future of this industry.

More recently there are new (emerging) applications based on the WAP and i-Mode that are driving the usage of mobile data networks that we need to address briefly. In this chapter we first provide an overview of the mobile data services, then we provide detailed description of a couple of mobile data services to familiarize the reader with implementation of these systems, and finally we finish the chapter with a short description of WAP and i-Mode.

9.1.1 What Is Mobile Data?

By mobile data networks we refer to those services, technologies, and standards that are related to data services over wide area coverage areas spanning more than the local area or campus. Examples include metropolitan area wireless data services such as Metricom's Ricochet service and those that operate over the same coverage areas as cellular networks such as CDPD or GPRS. The history of evolution of mobile data services was provided in Table 1.2 and a comparative description of major alternatives in Table 1.7.

In addition to traditional mobile data services, discussed in Chapter 1 and summarized in Table 1.7, SMS services can also be considered a part of these systems. Short messaging services are embedded in digital cellular systems such as GSM. These services use the 10-digit keypad of the mobile terminal to type and display a message and use the digital cellular network facilities to deliver the message. In traditional mobile data services, the subscriber uses computer keyboards to enter the message. Considering this larger picture for mobile data services or wireless WANs, as shown in Figure 9.1 we can classify mobile data networks into three categories: independent, shared, and overlay networks based on the way they relate to the cellular infrastructure.

9.1.2 Independent Mobile Data

Independent networks have their own spectrum that is not coordinated with any other service and their own infrastructure that is not shared with any other service. These networks are divided into two groups according to the status of their operating frequency band. The first group uses independent spectrum in licensed bands. Examples of such networks are ARDIS and Mobitex, and historically they were the first mobile data services that were introduced. Such networks were not economically successful because the revenues generated, mostly from the vertical applications, could not justify the cost of the implementation of the infrastructure. For these networks to survive, either a sizable vertical market is needed, or the cost of implementation of the infrastructure should be reduced, or a horizontal killer application is needed. The TETRA network (see Table 1.7) is designed for public safety application that is a prosperous vertical market. It was defined by the ETSI

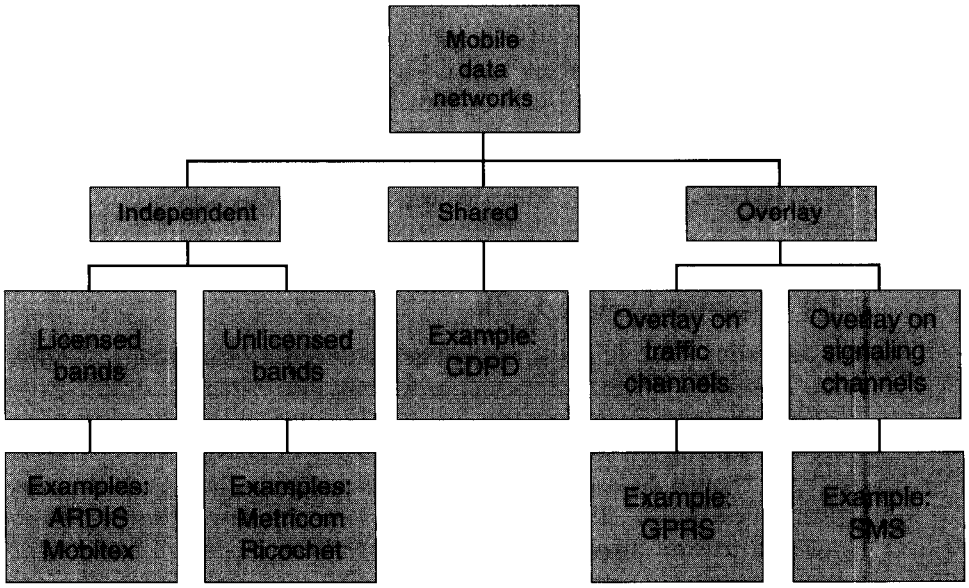

Figure 9.1 Classification of mobile data networks.

to meet the needs of professional mobile radio users. As a result, in spite of the fates of the previously mentioned mobile data projects, at the time of this writing, a number of companies in EU are engaged in manufacturing TETRA equipment. These companies have already reached agreements with the public safety organizations in Europe to use manufactured equipment in their future public land mobile data networks. Reduction of the cost of the infrastructure and killer applications are envisioned in the shared and overlay systems that are discussed in the following two sections.

The second group of independent mobile data networks makes use of unlicensed spectrum that is shared among a variety of applications and users. Metricom's Ricochet service, which used the 915 MHz unlicensed ISM band spectrum, was an example of this service. This service was deployed in airports and some metropolitan areas for wireless Internet access.

9.1.3 Shared Mobile Data

These networks share the spectrum and part of the infrastructure with an already existing voice-oriented analog service. The services operate in the same radio channels used for analog voice, but they have their own air-interface and MAC protocols. In addition to dedicated channels for data, these mobile data services can also use the available unused voice channels. These systems share an existing system infrastructure, therefore the initial investment is not huge, and it could be made as gradually as possible. Initial deployment could be made in areas where there is subscriber demand and subsequent penetration into other areas is considered as the customer base enlarges. The CDPD service (see Table 1.7 for overview), which shares spectrum and part of the infrastructure with AMPS, is an example of such networks. It does have an independent air-interface and MAC layer, along with additional infrastructure required for operation of data services. The CDPD standard was completed in the early 1990s, and the expectation was that by the year 2000, nearly 13 million subscribers would be using it. However, this expectation was never met, and generating income remained as the main obstacle for this service. Recently CDPD services have picked up with the availability of modems for palm computing devices. In the next section, we provide the details of implementation of CDPD to familiarize the reader with the implementation aspects of a data-oriented service.

9.1.4 Overlay Mobile Data

The last group of mobile data networks is an *overlay* on existing networks and services. This means that the data service will not only make use of the spectrum allocated for another service but also the MAC frames and air-interface of an existing voice-oriented digital cellular system. GPRS and GSM's SMS are examples of such overlays. They make use of free time slots available within the traffic channels and signaling channels in GSM. This way, the amount of new infrastructure required is reduced to a bare minimum. Most of the extra components required are implemented in software, making it inexpensive and easy to deploy. GPRS type of ser-

vices uses computer keyboards to communicate longer messages, and SMS use the cellular phone dialing keypad to communicate short messages. Sections 9.3 and 9.4 of this chapter provide further details on GPRS and SMS.

9.2 THE DATA-ORIENTED CDPD NETWORK

CDPD [TAY97], [CDPD95], [SAL99a,b] has been one of the longest surviving wide area mobile data technologies worldwide. It is a shared mobile data network that shares part of the infrastructure and the entire spectrum with AMPS in the United States. It is, however, an open standard, making its implementation easier and more widespread. CDPD initially had mixed success because the coverage was not universal, the data rates low, and prices prohibitive. CDPD has been more popular with vertical applications such as inventory for vending machines, fleet management, and so on. With the emergence of new handheld computers and palm-based devices, CDPD is making resurgence as a service for low data rate text-based Web, email, and short messaging horizontal applications. For example, modems are available for popular PalmOS and Windows CE devices that provide unlimited CDPD access.

In 1991, McCaw and IBM came up with the idea of developing a packet data overlay on AMPS. By the end of 1991, a telephone network–oriented prototype architecture was in place. This preliminary version was discussed in 1992 at a CDPD technical conference in Santa Clara by which point of time, CDPD also had the support of several regional telephone companies such as Ameritech, Bell Atlantic, GTE, McCaw, Nynex, PacTel, Southwest Bell, and US West. The first field trials were held in the San Francisco Bay area in 1992. At the same time, there was a recognition of the complexity of the telephony oriented architecture, and so open standards based on data networking standards and the OSI model were investigated around this time. Ultimately the latter approach based on data networking standards was selected, and the first official specifications were released in July 1993. The specifications were based on an open architecture, and the mobility management technique closely follows and is a precursor of the Mobile-IP standard. This specification was accepted by several of the major cellular telephony service providers. The CDPD forum was created in 1994, which attracted about 100 members. In 1995, release 1.1 of the CDPD standard came out. CDPD has been deployed by many of the former regional Bell operating companies as well as AT&T wireless.

9.2.1 What Is CDPD?

The design of CDPD was based on several design objectives that are often repeated in designing overlay networks or new networks. Some of these design goals are discussed. A lot of emphasis was laid on open architectures and reusing as much of the existing RF infrastructure as possible. The design goals of CDPD included location independence and independence from service provider, so that coverage could be

maximized; application transparency and multiprotocol support, interoperability between products from multiple vendors, minimal invention, and use of COTS technology as far as possible; and optimal usage of RF where air-interface efficiency is given priority over other resources. CDPD used primitive RF technology for cost reasons, and for this purpose the well-known GMSK modulation scheme was chosen. The raw signaling rate is 19.2 kbps, and with Reed-Solomon (RS) coding the effective data rate is 14.4 kbps full duplex before control overhead. The design was intended to be evolutionary and based on the OSI model with support for native IP, so that if new transport layer or application layer protocols were implemented there was no need for changes. The prominent features of CDPD have been its openness and freedom from all proprietary technology, support for multivendor interoperability, and simplicity in design. There is, however, constraint on the design of CDPD because it is a shared mobile data network and AMPS has priority over the usage of the spectrum. CDPD employs a technique called RF sniffing to detect whether an AMPS call is trying to access a frequency channel, and hopping to move from such a band to another to give the voice call priority. About 20 percent of AMPS frequency channels that can be used for CDPD are idle at a given time.

Example 9.1: Reuse of AMPS Infrastructure in CDPD

The AMPS infrastructure has a BS that transmits and receives RF signals. The demodulated voice signals are digitized and multiplexed on T1 connections that terminate in an MSC. The CDPD system aims at minimizing changes to existing infrastructure. In addition to the radio spectrum, the physical plant and communication links are precious resources. The MDBS (mobile data base station) (see following paragraph) handles RF communications and reuses existing antenna feeds and towers. The MDBS is a small compact box that can fit in existing cell sites and recent designs can be deployed in microcellular environments as well. The data is then relayed using *idle* channels in the T1 link to the MD-IS (mobile data intermediate system) that is usually colocated with the MSC.

Example 9.2: Channel Hopping in CDPD

When the telephone system selects a new channel for a voice call, if CDPD is using that channel it should exit within 40 ms. The MDBS should find an alternative channel and allocate it to the mobile end-system. The MDBS informs the mobile that the downlink is being changed to another channel and hops to this channel. A timed hop is a mandatory channel hop that is planned to happen after a fixed time. A forced hop is due to a voice call request. If no channels are available, CDPD enters "blackout."

9.2.1.1 CDPD Services

In a manner similar to our discussion of GSM, we start with the services that CDPD offers. *Network services* are the basic form of services offered by CDPD. This is simply support for transfer of data from one location to another via popular or standard network layer protocols. In particular, CDPD supports connectionless layer 3 protocols (IP or connectionless network protocol—CLNP) and in that sense

acts simply as a wireless extension to the Internet. *Network support services* are services necessary to maintain the operation of the mobile data network such as management, accounting, security, and so on. These services include mobility and radio resource management and are usually transparent to the user. Such services add "intelligence" to the network. The last category of services includes *network application services*. These are value added services such as limited size messaging on top of the network services and need explicit subscription.

9.2.2 Reference Architecture in CDPD

Figure 9.2 shows the reference architecture for CDPD. There are three key CDPD interfaces that form logical boundaries for a CDPD service provider's network. They are essential for the proper operation of CDPD. There are some interfaces internal to the "cloud." Such interfaces are only recommended, and a service provider can implement them differently. Each interface specifies a protocol stack corresponding to the OSI model and primitives are defined at each "layer" that can request and obtain services from the layer below.

9.2.2.1 Interface Details

The *A-Interface* is the air link interface, and parts 400–409 of the CDPD specifications specify it. The *E-Interface* is the external interface, and it is the means by which CDPD operates with the rest of data network. Over this interface IP and CLNP are supported, and IPv6 will be supported as it becomes deployed. Other protocols are supported by encapsulation because they are outside the CDPD specs. Mobility is transparent to the network beyond the E-interface. The *I-Interface* is the interservice provider interface. The North American market is partitioned into a multiplicity of service providers and the I-interface enables seamless nationwide service. It supports all of the E-interface protocols *plus* two CDPD specific protocols—the *mobile network location protocol* (MNLP) which is the

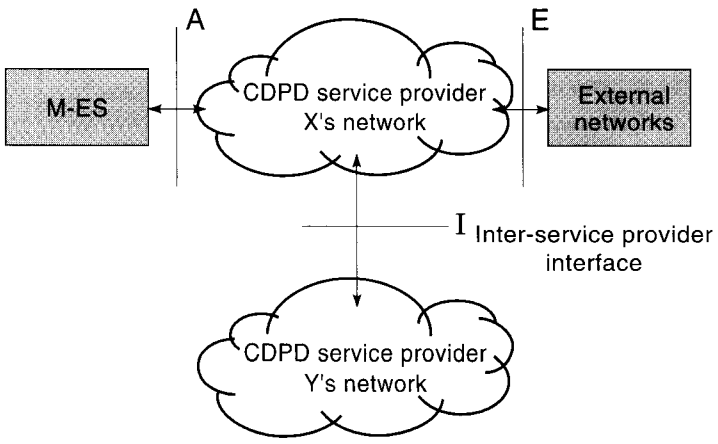

Figure 9.2 Reference architecture for CDPD.

protocol by which mobile users from one system are supported by another system and is a key piece of CDPD mobility management scheme. Network management and accounting protocols are also defined at the I-interface. All the protocols are based on CLNP (except reverse channel IP packets).

9.2.2.2 Physical Architecture

The physical elements of the CDPD architecture and their relationship with the three interfaces are shown in Figure 9.3. These elements are the mobile end system, the mobile data BS, and the mobile data intermediate system, along with some servers for accounting and network management and databases called the mobile home and mobile serving functions.

The mobile-end system (M-ES) is the ultimate source and destination of protocol data units (PDUs). It is equipped with a CDPD radio and software, and example M-ESs are telemetry devices, laptops, vending machines, and so on. In the M-ES, protocols are specified up to layer 3. An M-ES can be full duplex or half duplex and supports all standard APIs. Communication is implemented via conventional means like sockets, NDIS, and so on. An M-ES has three functional units: subscriber unit (SU), subscriber identity module (SIM), and mobile application subsystem (MAS). The subscriber unit establishes and maintains data communication, executes CDPD air-interface protocols, and includes administrative and management layers. The SIM is a repository of identity and authentication credentials, and the SIM card is very similar to GSM. In fact it is based on GSM standards. The MAS deals with the higher layer protocols and can perform remote database access, email, vending machine inventory, and so on. The MASs span a wide range of

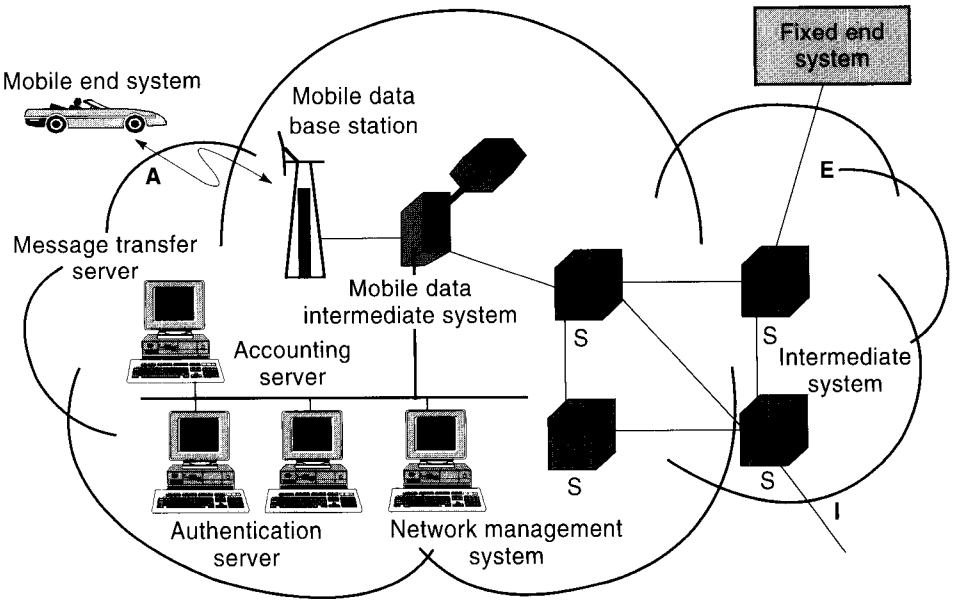

Figure 9.3 Physical elements in the CDPD architecture.

features. Some support both CDPD and an AMPS modem. Others can support voice communications as well. The M-ES employs a variety of power sources such as internal batteries, vehicular power source, or a laptop-based PC card.

The MDBS is the system end (network side) of the MAC sublayer over the air interface. It is equivalent to the BSS in GSM and is pretty much the BS electronics for CDPD. It is also like an Ethernet hub and acts as a link layer relay or bridge. It communicates with the M-ES through the A-interface and performs modulation of data bits and demodulation of the RF signal. It actively participates in the medium access scheme called digital sense multiple access (DSMA) (see Chapter 4). It is Layer 3 addressable for network management.

The MD-IS is the focal point of CDPD mobility management and packet forwarding. It has a mobile serving function (MSF) that serves as the foundation for registration of a mobile and a mobile home function (MHF) that serves as an anchor point for locating a mobile. It tracks the local point of access of the mobile devices and is responsible for presenting the external interface on behalf of the M-ES and for routing all traffic to and from the M-ES. It also performs accounting services. The MDBS to MD-IS interface is not specified though there are recommendations. Because it is internal to the CDPD cloud, proprietary implementations are possible although most service providers follow the recommendations. The intermediate system (IS) is a fancy name for a router, and it handles packet forwarding. Border ISs are further required to provide security filtering and access control functions that are *not* part of the CDPD specifications.

The fixed-end system (F-ES) is a conventional network node that includes most PCs, workstations, and so on that are transport layer peers of the M-ES. An Internal F-ES operates within the boundaries of the CDPD network and is under the control of the service provider. It usually operates functionally in the role of administrative servers and value-added servers. This is an example of the flexibility of CDPD.

Example 9.3: Internal F-ESs in CDPD

Internal F-ESs include an accounting server (AS), an authentication server, a directory server, a network management system, and a message transfer system. For example, the AS is in charge of collection and distribution of accounting data such as packet count, packet size, source and destination address, geographic information, time of transmission, and so on. CDPD employs pay by the byte (airlink usage) charging rather than time of connection. The network management system is based on general network management schemes of the OSI model using the common management interface protocol (CMIP), or optionally, simple network management protocol (SNMPv2).

9.2.3 Mobility Support in CDPD

As in the case of most mobile networks, mechanisms are in place in CDPD to support the mobile environment. We consider radio resources, mobility management, and security in this section.

9.2.3.1 Mobility Management

Handoff in CDPD occurs when an M-ES moves from one cell to another or if the CDPD channel quality deteriorates, the current CDPD channel is requested by an AMPS voice call (forced hop), or the load on CDPD channels in the current cell is much more than the load on the channels in an overlapping cell.

The physical layer of CDPD provides the ability to tune to a specific RF channel, the ability to measure the received signal strength indication (RSSI) of the received signal, the ability to set the power of the M-ES transmitted signal to a specified level, and the ability to suspend and resume monitoring of RF channels in the M-ES. Both uplink and downlink channels are slotted. There is no contention on the downlink, and the MDBS will transmit link layer frames sequentially. On the uplink, a DSMA/CD (digital sense multiple-access with collision detection) protocol is employed. Collision detection is at the BS and informed to the MHs on the downlink. On the downlink, multiple *cell configuration messages* are broadcast, including for the given cell and its neighbors, the cell identifier, a reference channel for the cell, a value that provides the difference in power between the reference channel, and the actual CDPD data channel, a RSS bias to compare the RSS of the reference channels of the given cell and adjacent cells, and a list of channels allocated to CDPD within the given cell. RSS measurements are always done on the reference channel because the CDPD channel list may keep changing [CDPD95].

Upon powering on, the MH scans the air and locks on to the strongest "acceptable" CDPD channel stream it can find and *registers* with the mobile-data intermediate system (MD-IS) that serves the base station. This is done via the mobile network registration protocol (MNRP) whereby the MH announces its presence and also authenticates itself. Registration protects against fraud and enables CDPD network to know the mobile location and update its mobility databases. The MH continues to listen to the CDPD channel unless it (or the CDPD network) initiates a handoff.

CDPD mobility management is based on principles similar to mobile-IP. The details are shown in Figure 9.4. The MD-IS is the central element in the process. An MD-IS is logically separated into a home MD-IS and a serving MD-IS. A home MD-IS contains a subscription database for its geographic area. Each subscriber is registered in his home MD-IS associated with his home area. The IP address of a subscriber points to his home MD-IS. At the home MD-IS, an MHF maintains information about the current location of MHs associated with (homed at) that home MD-IS. The MHF also encapsulates any packet that is addressed to an M-ES homed with it directing it to an MSF associated with the serving MD-IS, whose serving area the M-ES is currently visiting. A serving MD-IS manages one serving area. Mobile data BSs that provide coverage in this area are connected to the serving MD-IS, whose MSF contains information about all subscribers currently visiting the area and registered with it. The MSF employs the mobile network location protocol (MNLP) to notify the MHF about the presence of the M-ES in its service area. The channel stream in which a subscriber is active is also indicated. The MSF decapsulates forwarded packets and routes them to the correct channel stream in the cell.

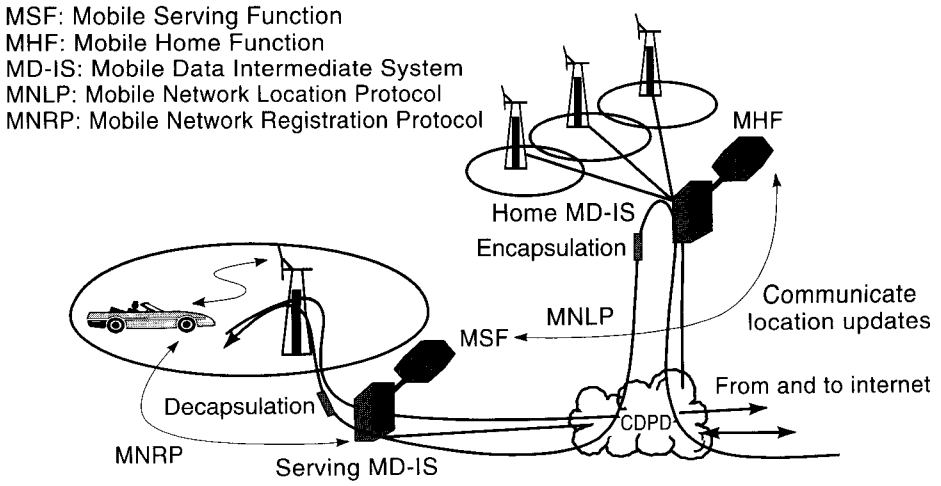

Figure 9.4 Mobility management in CDPD.

The handoff procedure in CDPD is shown in Figure 9.5. The handoff initiation and decision in CDPD are as follows. The handoff is mobile controlled. The M-ES always measures the signal strength of the reference channel [SAL99b]. An M-ES scans for alternative channels when its signal deteriorates. Because certain cells may have large shadowing effects within them, the operator can set a RSSI scan value to determine when a M-ES should start scanning for alternative channels. An M-ES will ignore a drop in signal level if the RSSI scan value is large enough or start scanning

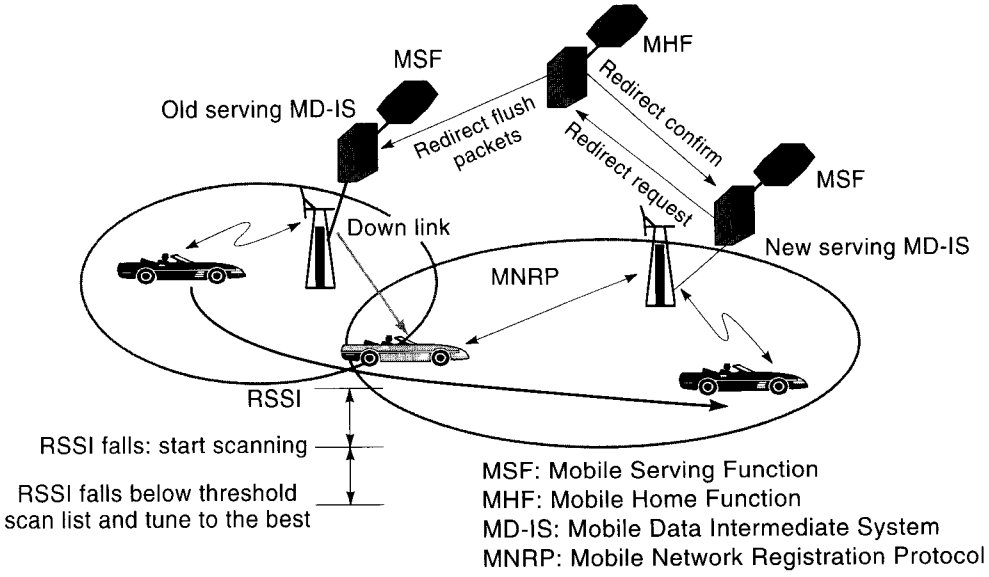

Figure 9.5 Handoff procedure in CDPD.

for alternative channels if it is small. This value is also useful (and should be made small) when the signal strength does not drop even when the M-ES has moved well into a neighboring cell. When additional thresholds for RSSI hysteresis, block error rate (BLER) and/or symbol error rate (SER) are reached, the M-ES will go through a list of channels of adjacent cells that the current BS is broadcasting and tune in to the one with the best signal strength. The M-ES informs the new BS that it has entered its cell. The mobile serving function of the new MD-IS uses a redirect-request and redirect-confirm procedure with the mobile home function of M-ES. The mobile home function also informs the old serving MD-IS about the handoff and directs it through its MSF to redirect packets it may have received for the M-ES to the new serving MD-IS or flush them. Depending on the nature of handoff (interoperator or intraoperator), the delay of registration and traffic redirection will vary.

9.2.3.2 Security

The security functions in CDPD are limited to data link confidentiality and M-ES authentication. There are some mechanisms for key management, the ability to upgrade, access control based on location, a network entity identifier (NEI), and screening lists.

CDPD authentication is performed by the *mobile network registration protocol management entity* (MME) that exists in both the MD-IS and M-ES. It uses the NEI along with an authentication sequence number (ASN) and an authentication random number (ARN) for authentication. The M-ES and network both maintain two sets of ARN/ASN-tuples (in case a fresh ARN is lost due to poor radio coverage). The shared historical record is the basis for authentication. It is updated every 24 hours. Authentication may be initiated at any time by the network.

CDPD confidentiality is based on encrypting all data using a secret key that is different in each session. The usual concept of using public key for exchanging keys and secret keys for block data encryption is employed. The session key generation is based on "Diffie-Hellman" key exchange with 256 bit values. It is, however, susceptible to the man-in-the-middle attack. RC-4 is used for block data encryption, and it is not a very secure secret key algorithm. Consequently, security limitations exist in CDPD. The CDPD network is not authenticated to the mobile as masquerading as a CDPD network is assumed to be impossible. Data confidentiality is not end-to-end, and it is assumed to be a higher layer issue. There are no mechanisms for data integrity, nonrepudiation, or traffic flow confidentiality.

9.2.3.3 Radio Resource Management

RRM is handled by a management layer in CDPD [SAL99b] and contains the procedures to handle the dynamically changing RF environment. In particular it takes care of (1) acquiring and releasing channels due to competition between CDPD and AMPS, and (2) handoff from one cell to another or from one channel to another. Its function is to continuously provide the best possible RF channel between the M-ES and the fixed network. The procedures are distributed between the M-ES and the network. The elements, algorithms, and procedures reside in the MDBS on the network side. The RRM also ensures that transmission power levels are set dynamically to minimize cochannel interference and optimize communication quality on the reverse

channel. In the MDBS, the RRM handles distribution of network configuration data, and power control data help the M-ES to track channel hops, perform handoffs, and satisfy power control requirements. On the M-ES side, the RRM has the algorithms and procedures to acquire and track CDPD forward channel transmissions, maintain the best possible CDPD channel in the area where the M-ES is located, and keep transmission power at the required level. All these may be assisted by data transmitted by the network. Unlike voice networks, the M-ES is supposed to handle RRM because the nature of transmission is extremely bursty. As such, the RRM functionality can be provided *with* or *without* the assistance of data provided by the network.

Example 9.4: Messaging for Power Control in CDPD

The channel stream identification (CSI) message (see Figure 9.6) on the downlink provides information about the current channel and also contains the following parameters: a cell identifier, a channel stream identifier, the service provider identity (SPI), a wide area service identifier (WASI), and finally a power product and maximum power level. The transmission power of the M-ES is calculated via the formula [SAL00]:

$$\text{Transmission power (dBW)} = \text{Power product (dB)} - 143\ \text{dBW} - \text{RSSI (dBW)}$$

RSSI is the received signal strength indication that is calculated from the received signal.

9.2.4 Protocol Layers in CDPD

The CDPD standard specifies a protocol stack for CDPD as shown in Figure 9.7. There are four layers: the physical layer, the MAC layer, the data link layer, and the subnetwork dependent convergence protocol (SNDCP) layer.

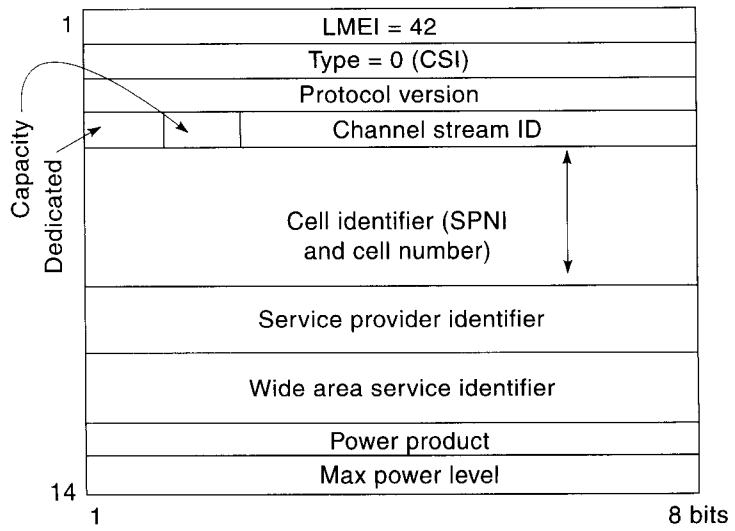

Figure 9.6 The channel stream identification message format.

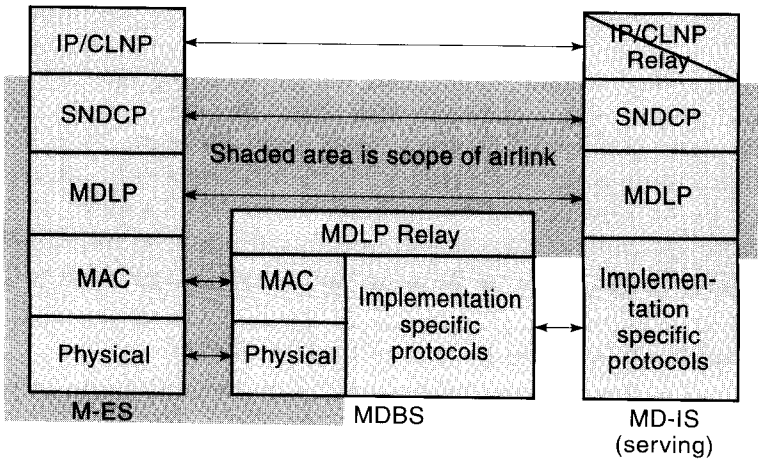

Figure 9.7 The CDPD protocol architecture.

The *physical layer* specifies two distinct one-way RF channels. The *forward channel* is from the MDBS to the M-ES and the *reverse channel* is from the M-ES to the MDBS. They are shown in Figures 9.8 and 9.9, respectively. Each channel is 30 kHz wide and corresponds to the same frequencies as AMPS. The transmission is digital GMSK at 19.2 kbps.

Example 9.5: Error Control Coding in CDPD

The physical layer is robust by employing a (63,47) RS coding. On the forward channel, one RS-block is transmitted every 21.875 ms, so that the raw bit rate is 420 bits/21.875 ms = 19.2 kbps. Up to eight symbol errors can be corrected with this code, but the CDPD specifications suggest correcting only up to seven symbol errors. The undetected symbol error rate is 2.75×10^{-8}.

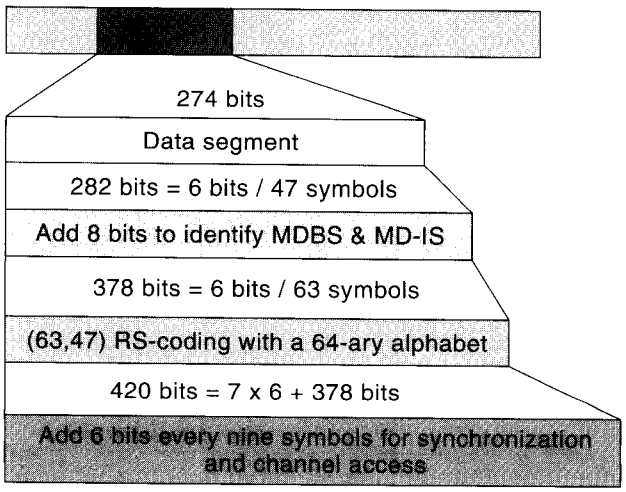

Figure 9.8 The forward channel in CDPD.

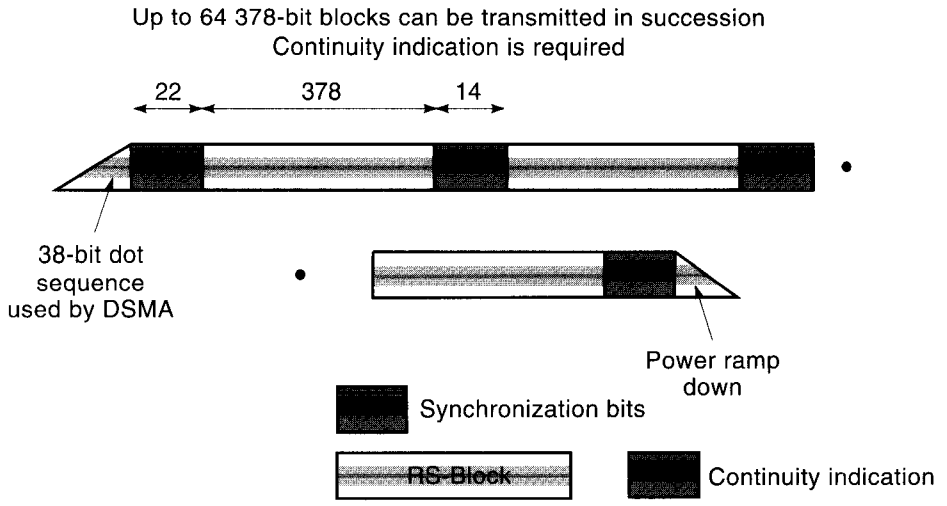

Figure 9.9 The reverse channel in CDPD.

The MAC layer follows the familiar CSMA/CD scheme typified by Ethernet. The forward channel has only one transmitter, the MDBS, and there is no contention for this channel. The reverse channel has multiple M-ESs competing for access. An M-ES may transmit on a channel whenever it has data to send and the channel is *not* already occupied by another transmission. The following steps are followed: (1) An M-ES must assess the state of the channel to see if it is available; (2) If the channel is occupied, there is a random back-off period (nonpersistent); (3) The M-ES transmits if the channel is free; and (4) If two M-ESs find the channel free and transmit simultaneously, there is a collision that must be detected to ensure retransmission. On a wireless link, there are problems with collision detection (see Chapter 4). Carrier sensing is easier at baseband than at RF frequencies. The receive and transmit frequencies are different. Also, transmissions from ground level can be detected at a tower but not at the ground level. Circuitry cost and power consumption become prohibitive for collision detection by a M-ES, so the channel-busy status and collision detection is enabled by the MDBS. A digital indicator is provided on the forward channel in order to indicate reverse channel status. This indicator is called the BUSY/IDLE indicator and it is set to BUSY whenever the MDBS senses reverse channel transmissions. The channels are slotted so that transmissions occur only in time slots. Another flag on the forward channel called *decode status* indicates whether the transmission was successfully decoded. Five bits are used to indicate busy/idle status and M-ESs use a simple majority to decide whether the channel is busy or idle.

Example 9.6: Flexibility in the MAC Protocol

The MAC protocol specifies a Min_idle_Time that is the minimum amount of time a mobile must wait after one burst to transmit again to ensure fairness. Two counts Min_count and Max_count specify limits of back-off delay periods in the case where there are too many or too few users. For example, if Min_count = 4, on the

first observation that the channel is busy or the first collision, the mobile backs off for a random number uniformly distributed between zero and $2^4 = 16$ slots. On a consecutive observation of busy channel, the distribution changes to twice the number of slots, for instance, random in (0,32) slots. The value Max_count = 8 implies, the maximum distribution interval is $2^8 = 256$ slots. Max_blocks is the maximum number of RS-blocks that can be transmitted in a burst. It's default value is 64, which implies that at most a 2 kbyte packet can be sent.

The *data link layer* uses the mobile data link protocol (MDLP) to connect the MD-IS and the M-ES. This is at the LLC (logical link control) layer equivalent of the 802.3 protocol. The "logical link" is identified by a "temporary equipment identifier" (TEI). The TEI reduces load on the air-interface and ensures privacy of the user. MDLP is similar to the LAPD of ISDN, and it uses a sliding window protocol. A selective reject mechanism is used. The sender has to retransmit only that frame that is acknowledged by an SREJ message. There is, however, no CRC check as in LAPD and the error control is shifted to the MAC layer.

Example 9.7: Other Data Link Features

Zap frames are used by the MD-IS to disable transmissions from a mobile for a given period of time. This is in case a M-ES does not follow the CDPD protocol in backing off or is not correctly working. A sleep mode is also available to power down if not transmitting to save battery life. The link layer is maintained in suspended mode, and all timers are saved. If a mobile does not transmit for a time period called T203 it is assumed to have gone to sleep. T203 is used to determine the "wake-up" time. Every T204 seconds the network broadcasts a list of TEIs that have outstanding data packets to receive. Before sleeping the mobile tracks the last T204 broadcast and wakes up at the next T204 to listen to the broadcast. After N203 TEI broadcasts, if a mobile is not up, the network discards the packets. Packet forwarding is also possible by storing it in the network for future transmissions.

The SNDCP maps MDLP services to those expected by IP or CLNP. It manages the difference between maximum data link frame size of 130 bytes, network packet size of up to 2048 bytes, and multiple network connections using the same MDLP. The functions of SNDCP include segmentation and reassembly, multiplexing, header compression, TCP/IP header compression using the Van Jacobson method, and the CLNP header compression using similar process and data encryption. Details of CDPD protocols, message formats, and so on can be found in [TAY97].

9.3 GPRS AND HIGHER DATA RATES

GPRS is an overlay on top of the GSM physical layer and network entities. It extends data capabilities of GSM and provides connection to external packet data networks through the GSM infrastructure with short access time to the network

for independent short packets (500–1,000 bytes). There are no hardware changes to the BTS/BSC (compared with CDPD), easy to scale, support for voice/data and data only terminals, high throughput (up to 21.4 kbps), and user-friendly billing.

9.3.1 What Is GPRS?

GPRS [KAL00], [CAI97], [BRA97] is an enhancement of the GSM. It uses exactly the same physical radio channels as GSM, and only new logical GPRS radio channels are defined. Allocation of these channels is flexible: from one to eight radio interface timeslots can be allocated per TDMA frame. The active users share timeslots, and uplink and downlink are allocated separately. Physical channels are taken from the common pool of available channels in the cell. Allocation to circuit switched services and GPRS is done dynamically according to a "capacity on demand" principle. This means that the capacity allocation for GPRS is based on the actual need for packet transfers. GPRS does not require permanently allocated physical channels. GPRS offers permanent connections to the Internet with volume based charging that enables a user to obtain a less expensive connection to the Internet. The GPRS MSs (terminals) are of three types. Class A terminals operate GPRS and other GSM services simultaneously. Class B terminals can monitor all services, but operate either GPRS or another service, such as GSM, one at a time. Class C terminals operate only GPRS service. This way there are options to have high-end or low-end terminals.

GPRS has some limitations in that there is only a limited cell capacity for all users and speeds much lower in reality. There is no store and forward service in case the MS is not available. The more popular short messaging service provides this feature, as we shall see in the next section.

The adaptation of GPRS to the IS-136 TDMA cellular standard is called GPRS-136. It is very similar to GPRS except that it uses 30 kHz physical channels instead of 200 kHz physical channels. Also there is no separate BSC. It can use coherent 8-PSK in addition to $\pi/4$-DQPSK to increase throughput over a limited area. This concept is similar to the 2.5G data service called *enhanced data rates for global evolution* (EDGE). Hooks in the standard allow the possibility of 16-QAM, 16-PSK, or 16-DQPSK in the future [SAR00].

9.3.1.1 GPRS Network Services

GPRS provides the following network services—point-to-multipoint (PTM-M) that is a multicast service to all subscribers in a given area, point-to-multipoint (PTM-G) that is a multicast service to predetermined group that may be dispersed over a geographic area, and point-to-point (PTP) service which is packet data transfer. This is of two types: connectionless based on IP and CLNP called PTP-CLNS and connection-oriented based on X.25 (PTP-CONS). GPRS also provides a bearer service for GSM's SMS discussed later in this chapter. There is also an anonymous access for MS at no charge. This is for example similar to an 800 number service where an agency that charges toll could allow an MS to access its credit card verification service for free.

GPRS has parameters that specify a QoS based on service precedence, a priority of a service in relation to another service (high, normal, and low), reliability,

and transmission characteristics required. Three reliability cases are defined and four delay classes. Here delay is defined as the end-to-end delay between two MSs or between an MS and the interface to the network external to GPRS. The reliability and delay classes are outlined in Tables 9.1 and 9.2. Transmission characteristics are specified by the maximum and mean bit rates. The maximum bit rate value can be between 8 kbps and 11 Mbps. The mean bit rate value is 0.22 bps to 111 kbps.

9.3.2 Reference Architecture in GPRS

As already mentioned, GPRS reuses the GSM architecture to a very large extent. There are a few new network entities called GPRS support nodes (GSN) that are responsible for delivery and routing of data packets between the mobile station and external packet network. There are two types of GSNs, the *serving GPRS support node* (SGSN) and the *Gateway GPRS support node* (GGSN). These are comparable to the MD-IS in CDPD. There is also a new database called the GPRS register (GR) that is colocated with the HLR. It stores routing information and maps the IMSI to a PDN address (IP address for example). Figure 9.10 shows this reference architecture.

The U_m interface is the air-interface and connects the MS to the BSS. The interface between the BSS and the SGSN is called the G_b interface and that between the SGSN and the GGSN is called the G_n interface. The SGSN is a router that is similar to the foreign agent in Mobile-IP. It controls access to MSs that may be attached to a group of BSCs. This is called a *routing area* (RA) or *service area* of the SGSN. The SGSN is responsible for delivery of packets to the MS in its service area and from the MS to the Internet. It also performs the logical link management, authentication, and charging functions. The GGSN acts as a logical interface to the Internet. It maintains routing information related to a MS, so that it can route packets to the SGSN servicing the MS. It analyses the PDN address of the MS and converts it to the corresponding IMSI and is equivalent to the HA in Mobile-IP.

9.3.3 Mobility Support in GPRS

In a manner similar to GSM and CDPD, there are mechanisms in GPRS to support mobility. We will discuss these issues in the following sections.

Table 9.1 Reliability Classes

	Probability for			
Class	***Lost Packet***	***Dupilcated Packet***	***Out-of-Sequence Packet***	***Corrupted Packet***
1	10^{-9}	10^{-9}	10^{-9}	10^{-9}
2	10^{-4}	10^{-5}	10^{-5}	10^{-6}
3	10^{-2}	10^{-5}	10^{-5}	10^{-2}

Table 9.2 Delay Classes

	128 Byte Packet		1,024 Byte Packet	
Class	***Mean Delay***	***95% Delay***	***Mean Delay***	***95% Delay***
1	< 0.5s	< 1.5s	< 2s	< 7s
2	< 5s	< 25s	< 15s	< 75s
3	< 50s	< 250s	< 75s	< 375s
4	Best Effort	Best Effort	Best Effort	Best Effort

9.3.3.1 Attachment Procedure

Before accessing GPRS services, the MS must register with the GPRS network and become "known" to the PDN. The MS performs an attachment procedure with an SGSN that includes authentication (by checking with the GR). The MS is allocated a temporary logical link identity (TLLI) by the SGSN and a PDP (packet data protocol) context is created for the MS. The PDP context is a set of parameters

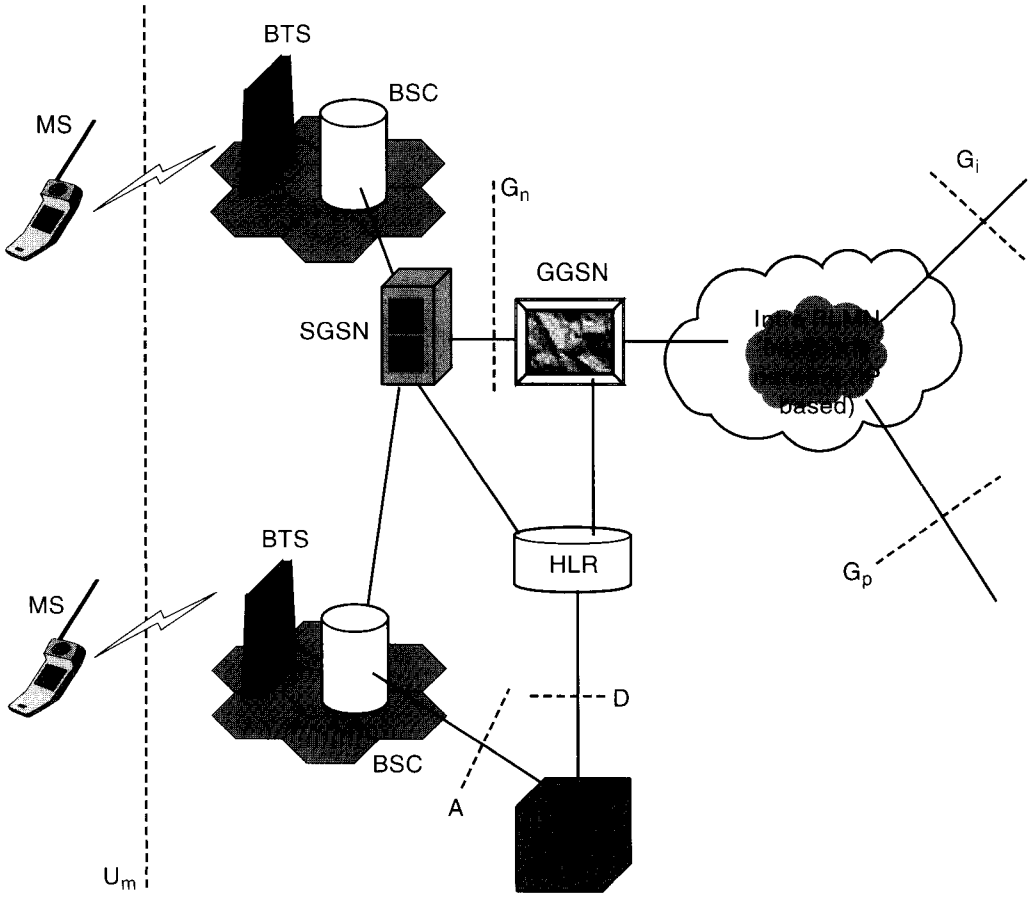

Figure 9.10 GPRS system architecture.

created for each session and contains the PDP type, such as IPv4, the PDP address assigned to the MS, the requested QoS parameters, and the GGSN address that serves the point of access to the PDN. The PDP context is stored in the MS, the SGSN, and the GGSN. A user may have several PDP contexts enabled at a time. The PDP address may be statically or dynamically assigned (static address is the common situation). The PDP context is used to route packets accordingly.

9.3.3.2 Location and Handoff Management

The location and mobility management procedures in GPRS are based on keeping track of the MSs location and having the ability to route packets to it accordingly. The SGSN and the GGSN play the role of foreign and HAs (visiting and home databases) in GPRS.

Location management depends on three states in which the MS can be (Figure 9.11). In the IDLE state the MS is not reachable, and all PDP contexts are deleted. In the STANDBY state, movement across routing areas is updated to the SGSN but not across cells. In the READY state, every movement of the MS is indicated to the SGSN. The reason for the three states is based on discussions similar to those in Chapter 6. If the MS updates its location too often, it consumes battery power and wastes the air-interface resources. If it updates too infrequently, a systemwide paging is needed; this is also a waste of resources. A standby state focuses the area to the service area of the SGSN. In the standby state, there is a medium chance of packets addressed to the MS. The ready state pinpoints the area when the chances of packets reaching are high.

Routing area updates that are part of the standby state are of two types. In the intra-SGSN RA update, the SGSN already has the user profile and PDP context. A new temporary mobile subscriber identity is issued as part of routing area update "accept." The HLR need not be updated. In an inter-SGSN RA update, the new RA is serviced by a new SGSN. The new SGSN requests the old SGSN to send the PDP contexts of the MS. The new SGSN informs the home GGSN, the GR, and other GGSNs about the user's new routing context.

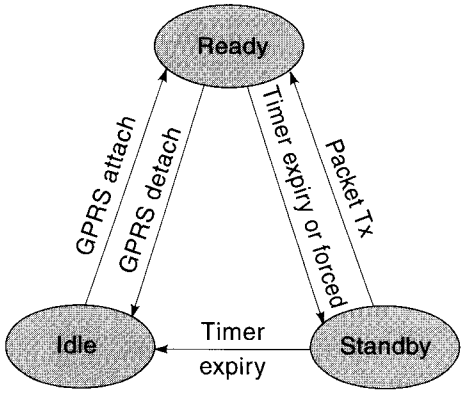

Figure 9.11 Location management in GPRS.

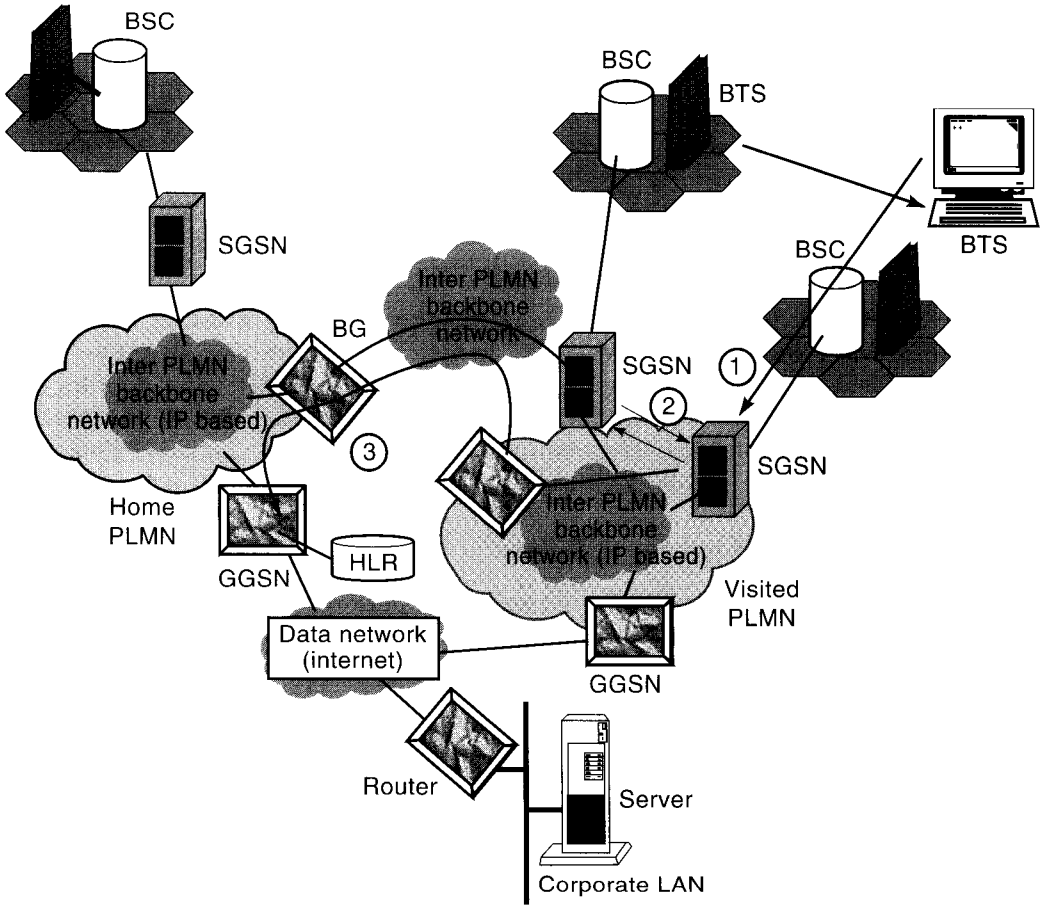

Figure 9.12 Handoff management in GPRS.

Mobility management in GPRS starts at handoff initiation. The MS listens to the BCCH and decides which cell it has to select. Proprietary algorithms are employed that use RSS, cell ranking, path loss, power budget, and so on. The MS is responsible for cell reselection independently, and this is done in the same way as GSM. The MS measures the RSS of the current BCCH and compares it with the RSS of the BCCH of adjacent cells and decides on which cell to attach it to. There is, however, an option available to operators to make the BSS ask reports from the MH (as in GSM), and then the handoff is done as in GSM (MAHO). Plain GPRS specific information can be sent in a *packet BCCH* (PBCCH), but the RSS is always measured from the BCCH. There are also other principles, which may be considered in handoff decision such as path loss, cell-ranking, and so on. The handoff procedure is very similar to mobile IP.

The location is updated with a routing update procedure, as shown in Figure 9.12. When a MS changes a routing area (RA), it sends an RA update request containing the cell identity and the identity of previous routing area, to the new SGSN (1). Note that an intra-SGSN routing area update (as discussed above) is also possible

when the same SGSN serves the new RA. The new SGSN asks the old SGSN to provide the routing context (GGSN address and tunneling information) of the MS (2). The new SGSN then updates the GGSN of the home network with the new SGSN address and new tunneling information (3). The new SGSN also updates the HLR. The HLR cancels the MS information context in the old SGSN and loads the subscriber data to the new SGSN. The new SGSN acknowledges the MS. The previous SGSN is requested to transmit undelivered data to the new SGSN.

9.4.3.3 Power Control and Security

Power control and security mechanisms are very similar to the way in which they are implemented in GSM (see Chapter 7). The ciphering algorithm is used to provide confidentiality and integrity protection of GPRS user data used for PTP mobile-originated and mobile-terminated data transmission and point-to-multipoint group (PTM-G) mobile terminated data transmission. The algorithm is restricted to the MS-SGSN encryption.

9.3.4 Protocol Layers in GPRS

In order to transport different network layer packets, GPRS specifies a protocol stack like CDPD and GSM (see Figure 9.13). This is the transport plane (where user data is transferred over the GPRS/GSM infrastructure). There is also a GPRS signaling plane to enable signaling between various elements in the architecture (like messaging between the SGSN and BSS etc.). GPRS employs out-of-band signaling in support of actual data transmission. Signaling between SGSN, HLR, VLR, and EIR is similar to GSM and extends only the GPRS related functionality. So it is based on SS-7. Between the MS and SGSN, a GPRS mobility management and session management (GMM/SM) protocol is used for signaling purposes.

The GPRS transport plane has different layers in different elements. The physical layers between the MS-BSS, BSS-SGSN, and SGSN-GGSN are indicated

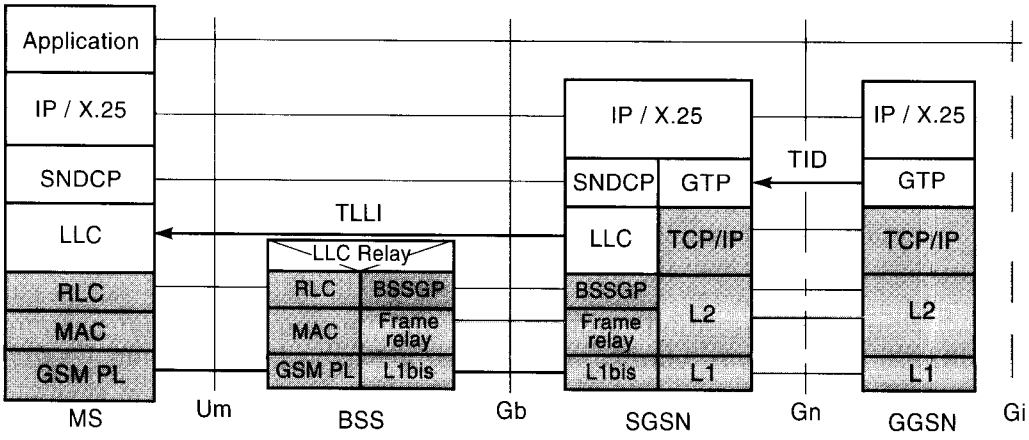

Figure 9.13 GPRS transport plane.

in Figure 9.13. Over the air, the physical layer is the same as GSM (i.e., it uses GMSK). Its functionalities include forward error correction and indication of uncorrectable code words, interleaving of radio "blocks," synchronization, monitoring of radio link signal quality, and so on. All other functions are similar to GSM. GPRS allows a MS to transmit on multiple time slots of the same TDMA frame unlike GSM. A very flexible channel allocation is possible since 1–8 time slots can be allocated per TDMA frame to a single MS. Uplink and downlink slots can be allocated differently to support asymmetric data traffic. Some channels may be allocated solely for GPRS. These are called packet data channels (PDCH).

Allocation of radio resources is also slightly different compared with GSM. A cell may or may not support GPRS and if it does support GPRS, radio resources are dynamically allocated between GSM and GPRS services. Any GPRS information is broadcast on the CCHs. PDCHs may be dynamically allocated or deallocated by the network (usually the BSC). If an MS is unaware that the PDCH has been deallocated, it may cause interference to a voice call. In such a case, fast release of PDCHs is achieved by a broadcast of a deallocation message on a PACCH.

The uplink and downlink transmissions are independent. The medium access protocol is called "Master-Slave Dynamic Rate Access" or MSDRA. Here, the organization of time-slot assignment is done centrally by the BSS. A "master" PDCH includes common control channels that carry the signaling information required to initiate packet transfer. The "slave" PDCH includes user data and dedicated signaling information for an MS. Several logical traffic and control GPRS channels are defined analogous to GSM. For example, PDTCH is the packet data traffic channel and PBCCH is the packet broadcast control channel. For random access to obtain a traffic channel, the packet random access, access grant, and paging channels are called PRACH, PAGCH, and PPCH, respectively. Additionally there is a packet notification channel (PNCH) that notifies arrival of a packet for the MS and a packet associated control channel (PACCH) used to send ACKs for received packets. A packet timing-advance control channel (PTCCH) is used for adaptive frame synchronization.

The packet transfer on the uplink and downlink are shown in Figures 9.14 and 9.15, respectively. They are quite similar to the process in GSM. Some of the differences are as follows. If a MS does not get an ACK, it will back off for a random time and try again. On the uplink, the Master-Slave mechanism utilizes a 3-bit uplink status flag (USF) on the downlink to indicate what PDCHs are available. A list of PDCHs and their USFs are specified in this USF. The packet resource or immediate assignment message indicates what USF state is reserved for the mobile on a PDCH. Channel assignment can also be done so that a MS can send packets uninterrupted for a predetermined amount of time. On the downlink, data transmission to a mobile can be interrupted if a high-priority message needs to be sent. Instead of paging, a resource assignment message may be sent to the MS if it is already in a "ready" state.

GPRS supports IP and X.25 packets at the network layer to be used by end-to-end applications. The SNDCP supports a variety of network protocols (IP, X.25, CLNP, etc.). All network layer packets share the same SNDCP. It multiplexes and demultiplexes the network layer payload and forms the interface between the link layer (LLC) and the network layer. Also the SNDCP handles packets based on QoS.

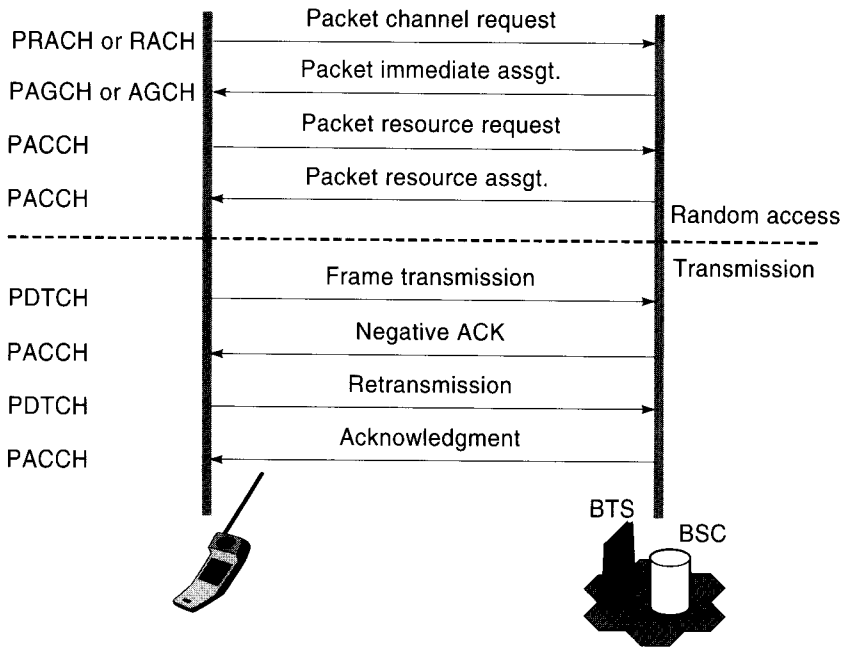

Figure 9.14 Uplink data transfer.

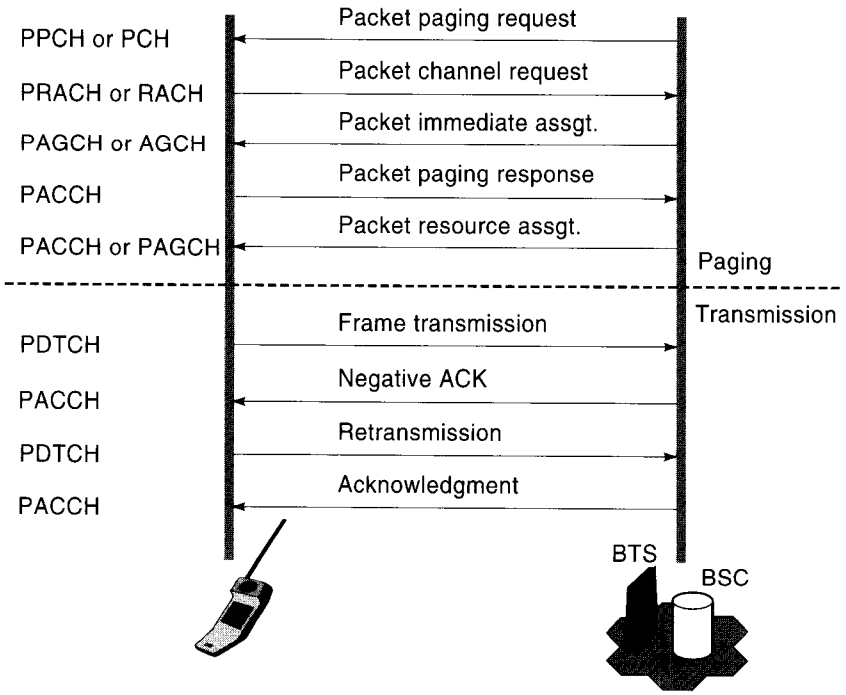

Figure 9.15 Downlink data transfer.

The LLC layer forms a logical link between the MS and the SGSN (similar to CDPD's MDLP). Each MS has a temporary logical link identity (TLLI) to identify itself in the LLC header. The LLC performs sequence control, error recovery, flow control, and encryption. It has an acknowledged mode (with retransmission for network layer payloads) and an unacknowledged mode (for signaling and SMS). The LLC also supports various QoS classes. Figure 9.16 shows how packets flow from higher layers, applications, and signaling levels to the SNDCP and the LLC. The packet transformation data flow is shown in Figure 9.17. The end result is blocks of 114 bits that are transmitted in bursts similar to GSM.

There are two levels of connections (tunneling mechanisms) implemented in the GPRS infrastructure as shown in Figures 9.13 and 9.18, one between the MS and the SGSN and the second between the SGSN and the GGSN. The two-level tunneling mechanism corresponds to a two-level mobility management: LLC "tunnels" (or virtual circuits) correspond to small area mobility, while GPRS tunneling protocol (GTP) tunnels correspond to wide area mobility. A new logical link is created each time the MS makes a handoff in the ready state between itself and the SGSN. If the SGSN does not change, the tunneling of the packet beyond the SGSN remains the same with the same GTP.

The BSS Gateway protocol (BSSGP) operates between the BSS and the SGSN relaying the LLC packets from the MS to the SGSN. Many MS LLCs can be multiplexed over one BSSGP. Its primary function is to relay radio related, QoS, and routing information between the BSS and SGSN and paging requests from SGSN to the BSS. It supports flushing of old messages from BSS. The data transfer is unconfirmed between BSS and SGSN.

The GTP allows multiprotocol packets to be tunneled through the GPRS backbone. A tunnel ID (TID) is created using the signaling plane that tracks the PDP context of each MS session. GTP can multiplex various payloads. The GTP

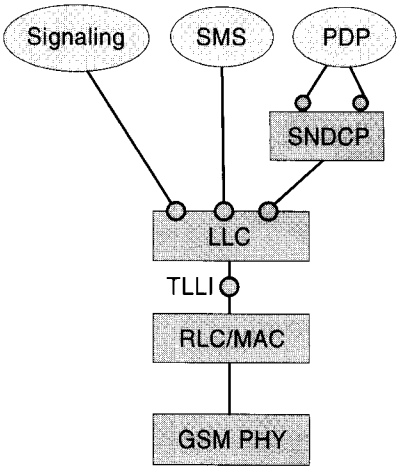

Figure 9.16 SNDCP and LLC in GPRS.

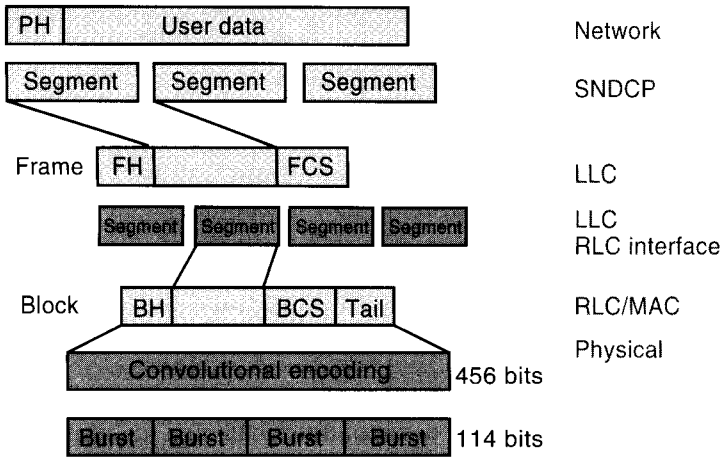

Figure 9.17 Packet transformation data flow.

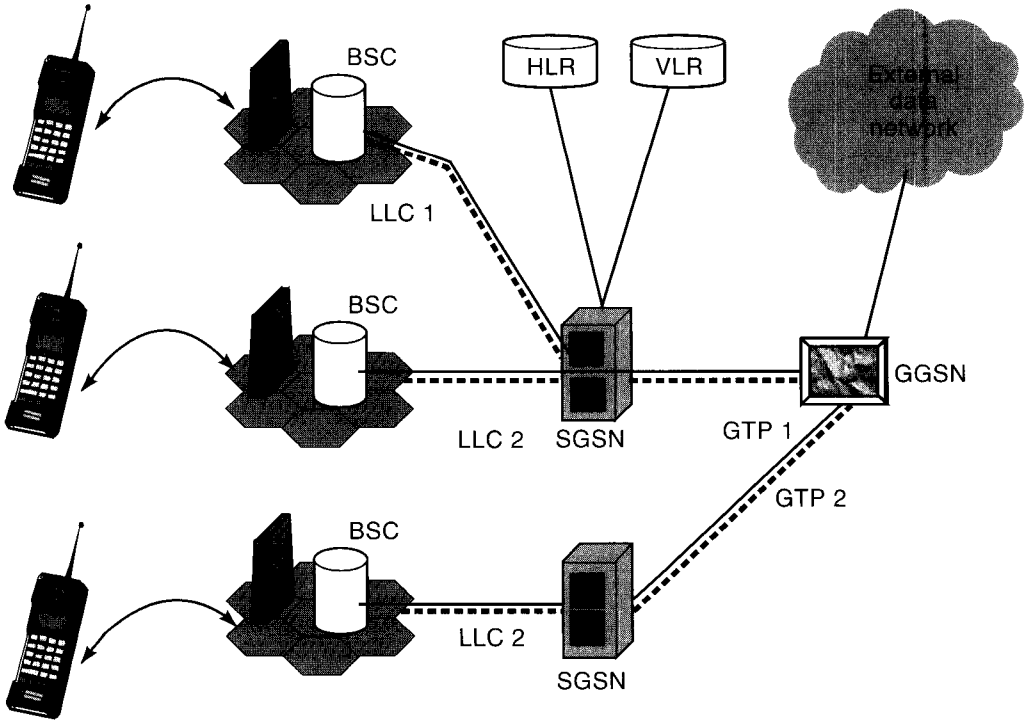

Figure 9.18 Two-level tunneling in GPRS.

packet is carried by either UDP/IP or TCP/IP, depending on whether the payload is IP or X.25, respectively.

9.4 SHORT MESSAGING SERVICES IN GPRS

The proliferation of GSM enabled the introduction of the *short messaging service* (SMS), which has become extremely popular in Europe. It is similar to the peer-to-peer instant messaging services on the Internet. Users of SMS [PEE00a], [PEE00b] can exchange alphanumeric messages of up to 160 characters (mapped into 140 bytes) within seconds of submission of the message. The service is available wherever GSM exists and makes it a very attractive wide area data service.

9.4.1 What Is SMS?

SMS was developed as part of GSM Phase 2 specifications (see Chapter 7). It operates over all GSM networks making use of the GSM infrastructure completely. It uses the same network entities (with the addition of a SMS center—SMSC), the same physical layer, and intelligently reuses the logical channels of the GSM system to transmit the very short alphanumeric messages.

9.4.1.1 Service Description

SMS has both an almost instant delivery service if the destination MS is active or a store and forward service if the MS is inactive. Two types of services are specified: In the *cell broadcast* service, the message is transmitted to all MSs that are active in a cell and that are subscribed to the service. This is an unconfirmed, one-way service used to send weather forecasts, stock quotes, and so on. In the *PTP* service, an MS may send a message to another MS using a handset keypad, a PDA, or a laptop connected to the handset, or by calling a paging center. Recently, SMS messages can be transmitted via dial-up to the service center and the Internet as well [PEE00a].

A short message (SM) can have a specified priority level, future delivery time, expiration time, or it might be one of several short predefined messages. A sender may request acknowledgment of message receipt. A recipient can manually acknowledge a message or may have predefined messages for acknowledgement.

An SM will be delivered and acknowledged whether a call is in progress because of the way logical channels in GSM are used for SMS. We discuss this in a later section.

9.4.2 Overview of SMS Operation

The SMS makes use of the GSM infrastructure, protocols, and the physical layer to manage the delivery of messages. Note that the service has a store-and-forward nature. As a result, each message is treated individually. Each message is maintained and transmitted by the SMSC. The SMSC sorts and routes the messages appropriately. The short messages are transmitted through the GSM infrastructure using SS-7. Details of the packet formats and messaging can be found in [PEE00a].

Figure 9.19 shows the reference architecture and the layered protocol architecture for SMS. There are two cases of short messages: a mobile originated SM and a mobile terminated short message. A SM originating from an MS has to be first delivered to a service center. Before that, it reaches an MSC for processing. A dedicated function in the MSC called the *SMS-interworking MSC* (SMS-IWMSC) allows the forwarding of the SM to the SMSC using a global SMSC ID. An SM that terminates at the MS is forwarded by the SMSC to the *SMS-gateway MSC* (SMS-GMSC) function in an MSC. As in the case of GSM, it either queries the HLR or sends it to the SMS-GMSC function at the home MSC of the recipient. Subsequently, the SM is forwarded to the appropriate MSC that has the responsibility of finally delivering the message to the MS. This delivery is performed by querying the VLR for details about the location of the MS, the BSC controlling the BTS providing coverage to the MS, and so on.

There are four layers in SMS—the application layer (AL), the transfer layer (TL), the relay layer (RL), and the link layer (LL). The AL can generate and display the alphanumeric message. The SMS-TL services the SMS-AL to exchange SMs and receive confirmation of receipt of SMs. It can obtain a delivery report or status of the SM sent in either direction. The RL relays the SMS PDUs through the LL. There are six PDUs types in SMS that convey the short message—from the SMSC to the MS and vice versa, convey a failure cause, and convey status reports and commands.

Over the air, the SMs are transmitted in time slots that are freed up in the control channels. If the MS is in an idle state (see Chapter 10), the short messages are sent over the SDCCH at 184 bits within approximately 240 ms. If the MS is in the active state (i.e., it is handling a call), the SDCCH is used for call set-up and maintenance. In that case, the SACCH has to be used for delivering the SM. This occurs at around 168 bits every 480 ms and is much slower. Failures can occur if there is a state change when the SM is in transit. The short message will have to be transmitted later.

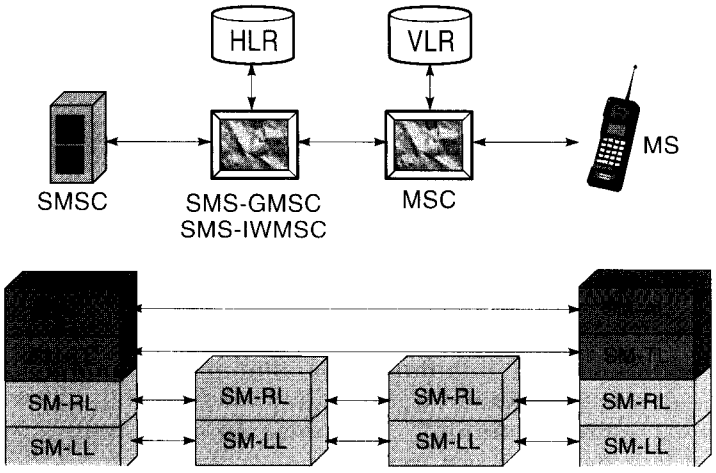

Figure 9.19 Reference and layered protocol architecture for SMS.

In the case of cell broadcast, a cell broadcast entity and a cell broadcast center are used to send the weather forecast or other broadcast SMs to multiple BSCs for delivery. The broadcasts contain the data and identities of MSs that are to receive the message. The cell broadcast is on the cell broadcast channel (CBCH).

9.5 MOBILE APPLICATION PROTOCOLS

Mobile applications are becoming very important in the age of the Internet. Data networks and dial-up services were mostly restricted to research laboratories and educational institutions in the early 1980s and became a household service by the late 1990s because of the emergence of applications such as email, e-commerce, and the World Wide Web. Recently, efforts provide similar applications on cellular networks. A major problem with providing such services on cellular networks has been the lack of resources such as bandwidth, processing power, memory, display sizes, interfaces like keypads, and so on that makes the service expensive. The constraints are also larger such as more latency, less connection stability, and less predictive availability. The wireless application protocol (WAP) and the i-mode service offered by NTT-DoCoMo of Japan are examples of how Internet based applications are being adapted to the cellular systems.

9.5.1 Wireless Application Protocol (WAP)

Initiated in 1997 by Nokia, Ericsson, Motorola and Phone.com, the WAP is an industry standard developed by the WAP Forum [WAPweb] for integrating cellular telephony and the Internet by providing Web content and advanced services to mobile telephone users. The WAP protocol is expected to help implementation of a variety of applications that include Internet access, m-commerce, multimedia email, tele-medicine, and mobile geo-positioning. The WAP application framework can run over several transport frameworks that include SMS, GPRS, CDPD, IS-136, and circuit switched wireless data services. Using WAP the wireless network technology will allow development of variety of applications. WAP orients the display toward text and material suitable for very small screens, thereby reducing both the load on the network and on the mobile terminal. In a nutshell, WAP attempts to optimize the Web and existing tools for a wireless environment.

WAP provides an extensible and scaleable platform for application development for mobile telephones. However, support for WAP on color PDA terminals is not very good. WAP also does not support seamless roaming between different link level bearer services such as CDPD and GSM, for example. Also, multimedia communications are not supported very well on WAP [FAS99]. For all these limitations, modifications to WAP will be necessary.

9.5.1.1 WAP Programming Architecture

How WAP does the extensions and modifications to the existing worldwide Web architecture is as follows. WAP introduces a *gateway* in between the wireless client and the rest of the Internet which manages the delivery of content to the mobile terminal. Such an architecture is shown in Figure 9.20.

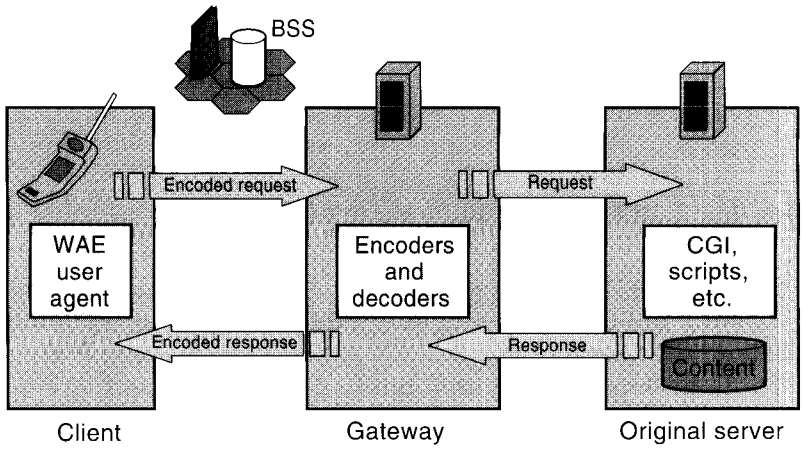

Figure 9.20 WAP programming architecture.

Most current Web content resides in Web servers and consists of a variety of material such as HTML pages, JavaScript interactive Web pages, images, and multimedia. These contents are developed for full-screen computer displays. A mobile terminal cannot access all this material because the screen size is smaller, the data rate for access is lower, and the cost of the access is higher than the wired access. To overcome this problem, the request for content is first made to the WAP Gateway, shown in Figure 9.20. The WAP request is made using a binary format of a *wireless markup language* (WML) that has some kind of a correspondence with HTML. That is, HTML pages can be converted into WML content. WML was derived from the extensible markup language (XML) and describes menu trees or decks through which a user can navigate with the microbrowser. The binary request conserves bandwidth by compressing the data. The request is decoded by the gateway and transmitted to the original server as an HTTP request. The server responds with the content in HTML. The content is filtered into WML, encoded into binary, and transmitted to the handset. A microbrowser in the handset coordinates the user interface and is similar to the usual Web browser. In essence, the gateway acts as a proxy device within the network. An example of a WAP network is shown in Figure 9.21. The WTA stands for wireless telephony application that provides an interface to a wireless application environment for network related activities such as call control, access to local address books, network events, and so on.

9.5.1.2 Protocol Layers in WAP

The layered protocol architecture specified for WAP is shown in Figure 9.22. The *wireless application environment* (WAE) is an application development platform that combines aspects of the Web and mobile telephony. It includes a microbrowser, WML, a scripting language similar to JavaScript called WMLscript,

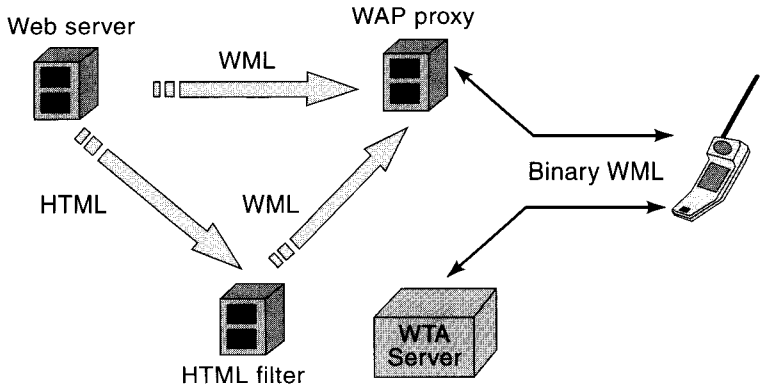

Figure 9.21 WAP example network.

several data formats such as images, calendars, and records, and telephony applications. The *wireless session protocol* (WSP) provides both a connection oriented and a connectionless service to the WAE. The connection-oriented service operates on top of the *wireless transport protocol* (WTP) and the connectionless service on top of the *wireless datagram protocol* (WDP). These are similar to Internet's TCP and UDP, respectively. WSP can suspend or migrate sessions unlike protocols on the wired link, and it is optimized for low bandwidth links.

The *wireless transport layer security* (WTLS) is based on the IETF standard transport layer security and the secure socket layer. It provides data integrity, privacy, authentication and techniques to reject replay attacks.

9.5.2 i-Mode

i-Mode is a service that tries to eliminate the use of a gateway and provide direct access to the Internet to the extent possible. With this goal in mind, Japan's NTT-DoCoMo has introduced this extremely popular service in Japan in 1999. By

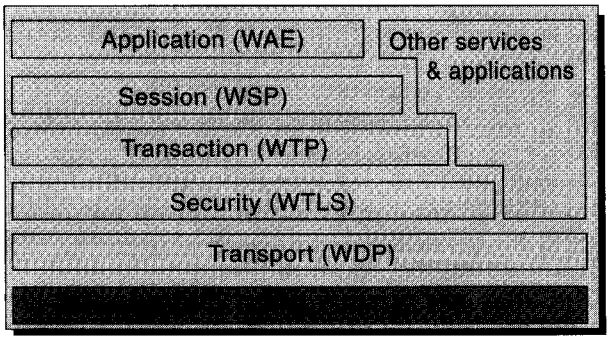

Figure 9.22 WAP layered protocol architecture.

November 2000 i-Mode had 14.9 million subscribers in Japan and Hong Kong. These terminals transmit data at 9,600 bps that allows graphics and small text messaging on a larger screen than the WAP. This display allows six to 10 lines of text at 16 to 20 characters per line that can be color or monochrome. i-Mode telephones can access HTML files across the web using C-HTML without a protocol like WAP. i-Mode is more similar to HTML which allows different computers to exchange information. i-Mode extends this to communicate with PDAs and i-Mode enabled cellular phones. One of the major features of i-Mode is that its charging mechanism is based on packet transmission rather than connection time which makes it less expensive for most users applications. The next version of WAP is expected to include CHTML as an alternative mechanism so as to include i-Mode as an option within itself.

QUESTIONS

9.1 What are the new elements added to the GSM infrastructure to support GPRS?
9.2 What are the new elements added to the AMPS infrastructure to support CDPD?
9.3 What are the duties of SGSN and GGSN in GPRS?
9.4 Why is the channel spacing in ARDIS and Tetra 25 kHz and in CDPD 30 kHz?
9.5 How does GPRS provides a variety of data rates?
9.6 How many classes of QoS are supported by GPRS and what are the differences between them?
9.7 Name the connectionless- and connection-oriented services provided by the GPRS.
9.8 What is the transmission band of the GPRS? Does it change when the data rate is changed?
9.9 What is the number of bits in each burst of GPRS and how does it differ from a GSM burst?
9.10 What are the expected latencies for circuit-switched and packet-switched data networks?
9.11 What is the difference between charging techniques for a packet data and a circuit-switched network?
9.12 With dedicated packet data networks such as ARDIS or MOBITEX, why would one choose CDPD?
9.13 What are the differences between GGSN and SGSN in GPRS, and MD-IS and MDBS in the CDPD?
9.14 What are the differences between the MAC layers of GPRS and CDPD?
9.15 What are the differences between the air interfaces of GPRS and CDPD?
9.16 What are the differences between the Mobile IP protocol and CDPD's method to support mobility?
9.17 What is the difference between full-duplex and half-duplex CDPD operations?
9.18 What is the maximum network layer throughput in CDPD for reverse and forward channels and why they are not the same?
9.19 What are the equivalents of CDPD's MHF and MSF in the Mobile IP protocol?
9.20 What are the MTP and PTP, and SMS?
9.21 Differentiate between the two types of independent mobile data networks.
9.22 Of the design goals of CDPD, which three do you consider important? Why?

9.23 Give three reasons why it is difficult to detect collisions at the transmitter in wireless networks.

9.24 What is the importance of color codes in CDPD?

9.25 Assume that you have a single mobile end system in a CDPD cell. What value of Min_Idle_Time would you suggest? Why?

9.26 Draw the protocol stack of CDPD at the M-ES, at the MDBS, and at the MD-IS. Show the communication between different peer layers.

9.27 What is GPRS-136? How does it differ from GPRS?

PROBLEMS

9.1 The fade rate of a channel is the number of times on average that the signal crosses an acceptable threshold. If the fade rate is larger than the packet rate, a packet cannot usually be received correctly over a wireless link. Packet sizes and data rates determine the critical fade rate that can be tolerated by a given system. Assume that there is no error correction or interleaving and the packet size in CDPD is 128 bytes at 19.2 kbps. If packets are continuously transmitted, what fade rate will destroy all packets? Suppose the data rate is increased to 56 kbps. How does the critical fade rate value change?

9.2 In CDPD, the M-ES has three base stations in its CCIB: BS1, BS2, and BS3. The quantities associated with these base stations are indicated by the subscripts 1, 2, and 3 in the following. The M-ES measures the RSS from a reference channel. To correct the measured value to reflect the actual RSS on the CDPD channel, it uses an ERP_Delta value, i.e., ERP_Delta = RSSI (reference channel) − RSSI (CDPD channel). An RSSI_bias value biases the handoff in favor of either the current cell or the neighboring cell. The RRME evaluates neighbor cells if

RSSI < PRSSI − RSSI_Scan_Delta for a time larger than RSSI average time *or*
RSSI > PRSSI + RSSI_Scan_Delta for a time larger than RSSI average time

This means the M-ES will start scanning for adjacent cells if the RSSI changes by RSSI_Scan_Delta for RSSI_Average_Time. This is called a non-critical condition. A critical condition implies for example that the BLER is higher than the threshold for a time larger than the BLER_average_time. The handoff procedure is as follows:

- Step 1: Close current channel. Compute RSSI_eff for each cell in the CCIB using:
  - RSSIeff = RSSI − ERP_delta + RSSI_Bias
  - Select best neighbor cells (with highest RSSI_eff)
- Step 2: If non-critical condition, compare current cell with best neighboring cells.
- If RSSIcurrent + RSSI_Hysteresis > RSSI_eff, stick with current cell else make a handoff.
- Skip Step 2 if a critical condition occurs.

Assume that the values recorded are as follows for a time larger than

RSSI_Average_Time. RSSI_Hysteresis = 3 dB

for the current base station (BS1).

RSSI 1 = −45 dBm	RSSI2 = −42 dBm	RSSI_bias2 = −6 dB
PRSSI1 = −50 dBm	RSSI3 = −57 dBm	RSSI_bias3 = 3 dB
BLER_Threshold = 10^{-3}	ERP_delta1 = 0 dB	ERP_delta2 = 0 dB
ERP_delta3=0 dB		

a. Explain what happens if the RSSI_Scan_delta is 6 dB and the measured BLER for a time greater than BLER_average_time is 10^{-4}? Why?

b. Explain what happens if the RSSI_Scan_delta is 4 dB and the measured BLER for a time greater than BLER_average_time is 10^{-4}? Why?

c. Explain what happens if the RSSI_Scan_delta is 6 dB and the measured BLER for a time greater than BLER_average_time is 1.5×10^{-3}? Why?

In each case explain whether a handoff is made, possibly to which base station, and why.

9.3 We know that the raw data rate at the physical layer in CDPD is 19.2 kbps. Calculate the actual data rate if you assume that the data is

a. before adding the control bits.

b. before performing the Reed-Solomon coding.

c. before frame delimiting.

What are your conclusions?

9.4 In CDPD, a (63,47) Reed-Solomon (RS) code is employed. The Reed-Solomon code encodes blocks of 47 symbols each carrying six bits into codewords of 63 symbols (also each of six bits). A codeword can correct up to seven symbol errors. A block of 63 symbols is transmitted roughly every 21 ms. What is the maximum fade duration that can be corrected by the RS code? How many codewords should be interleaved to combat fade durations of 105 ms? Explain how you arrive at your answers.

9.5 In 2.5G wireless data networks, sometimes called Enhanced Data rates for Global Evolution (EDGE), data throughput speeds of up to 384 kbps are proposed using existing GSM infrastructure. The idea with this approach is to use higher level modulation schemes like 8-PSK instead of GMSK or QPSK. This approach has moderate implementation complexity since the carrier spacing and TDMA frame structure remain the same. Only the channel coding and interleaving are different. Depending on the code rate and modulation scheme, the data rates provided are changed. Several combinations have been proposed for this purpose.

a. Determine the data rates that can be provided for the following cases:

i. CS-1 where GMSK is used with a code rate of 0.49

ii. CS-3 where GMSK is used with a code rate of 0.73

iii. PCS-3 where 8-PSK is used with a code rate of 0.6

iv. PCS-6 where 8-PSK is used with a code rate of 1

b. Discuss the implications of using the above combinations.

9.6 One of the problems with EDGE (see Problem 9.5) is the increased SNR requirement if higher level modulation schemes are employed. Assume that 8-PSK and QPSK are used without any coding in one implementation of EDGE. Using the Okumura-Hata model, determine what will be the coverage of a base station if it uses 8-PSK signal and if it uses QPSK. Assume that a bit error rate of 10^{-4} is required in either case. (*Hint:* Use the bit error rate formulas from Chapter 3. Assume that the transmit power is 30 dBm, the antenna heights are 75 m and 1 m, and the frequency of operation is 900 MHz.)