CHAPTER 7

Traffic Integration in Personal, Local, and Geographical Wireless Networks

RAFFAELE BRUNO, MARCO CONTI, and ENRICO GREGORI
CNR, Istituto CNUCE, Pisa, Italy

7.1 INTRODUCTION

Currently, users identify wireless networks with first- and second-generation cellular telephony networks. Although voice and short messaging have driven the success of these networks so far, data and more sophisticated applications are emerging as the future driving forces for the extensive deployment of new wireless technologies.

In this chapter, we will consider future wireless technologies that will provide support to different types of traffic including legacy voice applications, Internet data traffic, and sophisticated multimedia applications.

In the near future, wireless technologies will span from broadband wide-area technologies (such as satellite-based networks and cellular networks) to local and personal area networks. In this chapter, for each class of network, we will present the emerging wireless technologies for supporting service integration. Our overview will start by analyzing the Bluetooth technology [30] that is the de facto standard for wireless personal area networks (WPANs), i.e., networks that connect devices placed inside a circle with radius of 10 meters. Two main standards exist for wireless local area networks (WLANs): IEEE 802.11 [21] and HiperLAN [15]. In this chapter we focus on the IEEE 802.11 technology, as it is the technology currently available on the market. After a brief description of the IEEE 802.11 architecture, we will focus on the mechanisms that have been specifically designed to support delay-sensitive traffic.

For wireless wide area networks, we will focus on the technology for third-generation mobile radio networks. Two standards are emerging worldwide for this technology: the Universal Mobile Telecommunication System (UMTS) of the European Telecommunication Standard Institute (ETSI), and the International and Mobile Telecommunications-2000 (IMT-2000) of the International Telecommunication Union (ITU). The differences between these two standards are not relevant for the discussion in this chapter. Whenever necessary, we will use UMTS as the reference technology [1, 32].

All the network technologies analyzed in this chapter operate according to the infrastructure-based approach (see Figure 7.1). An infrastructure-based architecture imposes the existence of a centralized controller for each cell, which takes different names depend-
ing on the technology: master, access point, base station, etc. The cell identifies the area covered by the centralized controller, i.e., the area inside which a mobile terminal can directly communicate with the centralized controller. The cell size, as said before, depends on the technology, e.g., from 10 meters in Bluetooth up to kilometers in UMTS. Furthermore, inside UMTS, cells of different sizes can be used to accommodate different classes of users.

The centralized controller is connected to the wired network so as to have both intercell communication and access to other networks such as Internet.

WPANs and WLANs may also operate in the ad hoc mode [29]. An ad hoc network is a set of mobile terminals within the range of each other that dynamically configure themselves to set up a temporary network (see Figure 7.1). In this configuration, no fixed controller is required, but a controller is dynamically elected among all the stations participating in the communication.

Both in the infrastructure-based and ad hoc modes, the centralized controller is in charge to manage the radio resources of its cell. To achieve this, the following functionalities are implemented in all the network technologies we analyze: a medium access control mechanism, a scheduling algorithm, and a signaling channel for the communications from the centralized controller to the mobile terminals (downlink signaling channel).

The medium access control mechanism is required for managing the communications from the mobile terminals to the controller, and it is used by the mobile terminals for requesting transmission resources. In all technologies, this mechanism is used when a mobile terminal needs to start a communication and hence does not yet have any transmission resources allocated to it. In this case, the mobile terminal transmits on a channel that is shared among all the terminals in the cell. Protocols belonging to the random access class are typically used to implement the medium access control mechanisms [18]. Once the

Figure 7.1 Infrastructure-based and ad hoc networks.
centralized controller receives the mobile terminal requests, it assigns the transmission resources according to the rules defined by its scheduling algorithm. Finally, the assigned resources are communicated to the terminals through the downlink signaling channel.

As the emphasis of this chapter is on the integration of different types of traffic, we will primarily focus on the medium access control mechanisms, the scheduling algorithms, and the downlink signaling channels adopted by these technologies.

7.2 A TECHNOLOGY FOR WPAN: BLUETOOTH

Bluetooth wireless technology is a de facto standard for low-cost, short-range, radio links between mobile PCs, mobile phones, and other portable devices. The Bluetooth specifications are released by the Bluetooth Special Interest Group (SIG), an industry group consisting of industrial leaders in the telecommunications, computing, and networking [11]. In addition, the IEEE 802.15 Working Group for Wireless Personal Area Networks has started a project to publish and approve a standard derived from the Bluetooth specification [20].

The Bluetooth system operates in the 2.4 GHz industrial, scientific, and medicine (ISM) band. It is based on a low-cost, short-range radio link integrated into a microchip, enabling protected ad hoc connections for wireless communication of voice and data in stationary and mobile environments. It enables use of mobile data in different ways for different applications. Due to its low-cost target, it can be envisaged that Bluetooth microchips will be embedded in all consumer electronic devices.

The characteristics of the Bluetooth technology offer wide room for innovative solutions and applications that could bring radical changes to everyday life. Let us imagine a PDA (with a Bluetooth microchip) that automatically synchronizes with all the electronic devices in its 10 meter range when you arrive at your home. Your PDA can, for example, automatically unlock the door, turn on the house lights while you are getting in, and adjust the heat or air conditioning to your preset preferences. But not only the home can become a more comfortable environment when the access to information is fast and easy. Let us imagine arriving at the airport and finding a long queue at the check-in desk for seat assignment. You can avoid the queue using a hand-held device to present an electronic ticket and automatically select your seat.

7.2.1 The Bluetooth Network

From a logical standpoint, Bluetooth belongs to the contention-free, token-based multiaccess networks [18]. In a Bluetooth network, one station has the role of master and all other Bluetooth stations are slaves. The master decides which slave is the one to have access to the channel. The units that share the same channel (i.e., are synchronized to the same master) form a piconet, the fundamental building block of a Bluetooth network. A piconet has a gross bit rate of 1 Mbps that represents the channel capacity before considering the overhead introduced by the adopted protocols and polling scheme. A piconet contains a master station and up to seven active (i.e., participating in data exchange) slaves simultaneously. Independent piconets that have overlapping coverage areas may form a scatternet.
A scatternet exists when a unit is active in more than one piconet at the same time (a unit can be master in only one piconet). A slave may communicate with the different piconets it belongs to only in a time-multiplexing mode. This means that, for any time instant, a station can only transmit on the single piconet to which its clock is synchronized at that time. To transmit on another piconet, it has to change the synchronization parameters. More details on construction procedures for piconets and scatternets can be found in Chapter 27 of this handbook.

7.2.2 The Bluetooth Architecture

The complete protocol stack contains a Bluetooth core of Bluetooth-specific protocols: Bluetooth radio, baseband, link manager protocol (LMP), logical link control and adaptation protocol (L2CAP), service discovery protocol (SDP) as shown in Figure 7.2. In addition, examples of higher-layer non-Bluetooth-specific protocols are also shown in the figure; these can be implemented on top of the Bluetooth technology.

Bluetooth radio provides the physical links among Bluetooth devices and the baseband layer provides a transport service of packets on the physical links. In the next subsections these layers will be presented in detail.

The LMP protocol is responsible for the set-up and management of physical links. The management of physical links consists of several activities: putting a slave in a particular operating state (i.e., sniff, hold, or park modes [30]), monitoring the status of the physical channel, and assuring a prefixed quality of service (e.g., LMP defines transmission power, maximum poll interval, etc.). LMP also implements security capabilities at link level.

The radio, baseband, and LMP may be implemented in the Bluetooth device. The device will be attached to a host, thus providing that host with Bluetooth wireless communi-
cation. L2CAP layer and the other high-layer protocols are in the host. The host controller interface is a standard interface that enables high-layer protocols to access the services provided by the Bluetooth device.

The L2CAP services are used only for data transmissions. The main features supported by L2CAP are: protocol multiplexing (the L2CAP uses a protocol-type field to distinguish between upper-layer protocols) and segmentation and reassembly. The latter feature is required because the baseband packet size is smaller than the usual size of packets used by higher-layer protocols.

In legacy LANs, users locate services such as file server, print server, and name server by some static configuration. The configuration is usually established and maintained by a system administrator who manually configures the client devices. For dynamic ad hoc networks, this static configuration is not adequate. The SDP protocol is used to find the type of services that are available in the network.

Finally, RFCOMM is a serial line emulation protocol, i.e., a cable replacement protocol. It emulates RS-232 control and data signals over Bluetooth baseband, providing transport capabilities for upper-level services that use serial lines as their transport mechanism.

### 7.2.3 The Bluetooth Device

A Bluetooth unit consists of a radio unit operating in the 2.4 GHz band. In this band, 79 different radio frequency (RF) channels that are spaced 1 MHz apart are defined. The radio layer utilizes the frequency hopping spread spectrum (FHSS) as its transmission technique. The hopping sequence is a pseudorandom sequence of 79 hop length, and it is unique for each piconet. It is enabled by exploiting the actual value of the master clock and its unique Bluetooth device address, a 48 bit address compliant with the IEEE 802 standard addressing scheme [30]. The FHSS system has been chosen to reduce the interference of nearby systems operating in the same frequency range (for example, IEEE 802.11 WLAN) and make the link robust [12, 17]. The nominal rate of hopping between two consecutive RF is 1600 hop/sec.

A time division duplex (TDD) scheme of transmission is adopted. The channel is divided into time slots, each 625 μs in length, and each slot corresponds to a different RF hop frequency. The time slots are numbered according to the Bluetooth clock of the master. The master has to begin its transmissions in even-numbered time slots. Odd-numbered time slots are reserved for the beginning of the slaves’ transmissions.

The transmission of a packet nominally covers a single slot, but it may last up to five consecutive time slots (see Figure 7.3). For multislot packets, the RF hop frequency to be used for the entire packet is the RF hop frequency assigned to the time slot in which the transmission has begun. The RF change reduces the interference from signals coming from other radio modules.

There are two types of physical links that can be established between Bluetooth devices: a synchronous connection-oriented (SCO) link, and an asynchronous connectionless (ACL) link. The first type of physical link is a point-to-point, symmetric connection between the master and a specific slave. It is used to deliver delay-sensitive traffic, mainly voice. In fact, the SCO link rate is 64 Kbit/s and it is settled by reserving a couple of consecutive slots for master-to-slave transmission and immediate slave-to-master response.
The SCO link can be considered a circuit-switched connection between the master and the slave. The second kind of physical link, ACL, is a connection between the master and all slaves participating in the piconet. It can be considered a packet-switched connection between the Bluetooth devices and can support the reliable delivery of data: a fast automatic repeat request (ARQ) scheme is adopted to assure data integrity. An ACL channel supports point-to-multipoint transmissions from the master to the slaves.

As stated above, channel access is managed according to a polling scheme. The master decides which slave is the only one to have access to the channel by sending it a packet. The master packet may contain data or can simply be a polling packet. When the slave receives a packet from the master, it is authorized to transmit in the next time slot. For SCO links, the master periodically polls the corresponding slave. Polling is asynchronous for ACL links. Figure 7.4 presents a possible pattern of transmissions in a piconet with a master and two slaves. Slave 1 has both a SCO (packets filled with diagonal lines) and an ACL (packets filled with horizontal lines) link with the master, whereas Slave 2 has an ACL link only (packets filled with vertical lines). In this example, the SCO link is periodically polled by the master every six slots, whereas ACL links are polled asynchronously. Furthermore, the size of the packets on an ACL link is constrained by the presence of SCO links. For example, in Figure 7.4 the master sends a multislot packet to Slave 2, which, in turn, can reply with a single-slot packet only, because the successive slots are reserved for the SCO link.

As stated above, a piconet has a gross bit rate of 1 Mbps. The polling scheme and the protocols control information, obviously reducing the amount of user data that can be delivered by a piconet. We analyze the limiting performance of a piconet below. This analysis is performed by assuming a single master–slave link in which both stations operate under asymptotic conditions, i.e., the stations always have a packet ready for transmission. The results of this analysis are summarized in Tables 7.1 and 7.2 for SCO and ACL links, respectively. To enhance the reliable delivery of the packets, forward error correction (FEC) and cyclic redundancy check (CRC) algorithms may be used. The possible presence of FEC, CRC, and multislot transmission results in different payload lengths, as summarized in the tables.
The SCO packets (see Table 7.1), denoted by HV_y, are never retransmitted and the payload is not protected by a CRC. The y indicates the FEC level and it also identifies how many SCO connections may be concurrently active in a piconet. In addition to the three pure SCO packets, a DV packet is defined that can also carry asynchronous data but is still recognized on SCO links. In the Table 7.1, the items followed by “D” relate to the data field only. The ACL packets (see Table 7.2) are of two different groups, one denoted DM_x (medium-speed data) and the other one denoted DH_x (high-speed data). The former has a payload encoded with a 2/3 FEC and the latter has no FEC encoding. The subscript x

### TABLE 7.1 SCO packets

<table>
<thead>
<tr>
<th>Type</th>
<th>User payload (bytes)</th>
<th>FEC</th>
<th>CRC</th>
<th>Symmetric maximum rate (kbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HV1</td>
<td>10</td>
<td>1/3</td>
<td>no</td>
<td>64.0</td>
</tr>
<tr>
<td>HV2</td>
<td>20</td>
<td>2/3</td>
<td>no</td>
<td>64.0</td>
</tr>
<tr>
<td>HV3</td>
<td>30</td>
<td>no</td>
<td>no</td>
<td>64.0</td>
</tr>
<tr>
<td>DV</td>
<td>10 + (0–9)D</td>
<td>2/3 D</td>
<td>yes D</td>
<td>64.0 + 57.6 D</td>
</tr>
</tbody>
</table>

### TABLE 7.2 ACL packets

<table>
<thead>
<tr>
<th>Type</th>
<th>User payload (bytes)</th>
<th>FEC</th>
<th>CRC</th>
<th>Symmetric maximum rate (kbps)</th>
<th>Asymmetric maximum rate (kbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Forward</td>
<td>Reverse</td>
</tr>
<tr>
<td>DM1</td>
<td>0–17</td>
<td>2/3</td>
<td>yes</td>
<td>108.8</td>
<td>108.8</td>
</tr>
<tr>
<td>DM3</td>
<td>0–121</td>
<td>2/3</td>
<td>yes</td>
<td>258.1</td>
<td>387.2</td>
</tr>
<tr>
<td>DM5</td>
<td>0–224</td>
<td>2/3</td>
<td>yes</td>
<td>286.7</td>
<td>477.8</td>
</tr>
<tr>
<td>DH1</td>
<td>0–27</td>
<td>no</td>
<td>yes</td>
<td>172.8</td>
<td>172.8</td>
</tr>
<tr>
<td>DH3</td>
<td>0–183</td>
<td>no</td>
<td>yes</td>
<td>390.4</td>
<td>585.6</td>
</tr>
<tr>
<td>DH5</td>
<td>0–339</td>
<td>no</td>
<td>yes</td>
<td>433.9</td>
<td>723.2</td>
</tr>
</tbody>
</table>
stands for the number of slots that are necessary to transmit the packet. All ACL packets have a CRC field for checking the payload integrity. Tables 7.1 and 7.2 summarize SCO and ACL packet characteristics, respectively. In addition, the tables report, assuming a piconet with two only devices, the maximum aggregate piconet throughput for symmetric and asymmetric communications. In the asymmetric case, the throughput corresponding to DM\textsubscript{x} is computed by assuming that forward and the reverse traffic is transmitted using DM\textsubscript{x} and DM1 packets, respectively.

### 7.2.4 Scheduling Algorithms for the ACL Traffic

In the previous section, we examined the limiting performance of a Bluetooth piconet in the simple two-station configuration. In this configuration, Bluetooth is simply used as a cable replacement. However, as explained before, this technology is designed to operate in a more general piconet setting where there are several active slaves. In this case, the master must implement a scheduling algorithm to decide the slaves’ polling order. The Bluetooth specification indicates as a possible solution the round robin polling algorithm: slaves are polled in a cyclic order. Below, we evaluate Bluetooth performance via simulation, assuming a round robin scheduler. The simulated network topology is constituted by a single piconet with a master and six slaves.

We have modeled the intrapiconet communications, i.e., no traffic comes (goes) from (to) the outside of the piconet. Each slave is a source of IP packets and the interarrival times between consecutive packet generations are exponentially distributed, hence the IP packet arrival process is Poissonian. The packet length is uniformly distributed in the range from 500 to 1500 bytes. Each IP packet is encapsulated into an L2CAP packet that adds the 4 bytes L2CAP header and sent to the Bluetooth device local transmission queue. This local queue has a finite size $B_S$ and the queued packets are served according to a first come first served (FCFS) policy. Large L2CAP packets must be segmented into smaller baseband packets before transmission. A new L2CAP packet cannot be served until all fragments (generated during the segmentation) of the previous L2CAP packet have been successfully transmitted. The segmentation procedure is accomplished, just before the transmission, in such a way as to generate the minimum number of baseband packets.

Within the master, $N$ local transmission queues are implemented, where $N$ is the number of active slaves. Each master local queue has a finite size $B_M$ and the queued packets are served according to a FCFS policy. When an L2CAP packet is completely received by the master, the master accomplishes the reassembly procedure and forwards it on the transmission queue related to the slave, to which the packet is addressed.

In the transmission phase, the master behaves the same way as a slave. The master and the slaves transmit the ACL packets according to the Bluetooth transmission scheme described in the previous sections.

During the simulations we performed, we considered two traffic patterns: symmetric and asymmetric. In the former, all slaves contribute the same percentage to the offered load, while in the asymmetric case, Slave 1 produces the 90% of the overall load. In both traffic patterns, the destination address is sampled in a uniform way among the other slaves.

Simulative results presented in this section have been obtained by applying the independent replication technique with a 90% confidence level. Furthermore, we assumed an ideal
channel with no transmission errors [26]. Within each simulation, we have utilized the DH type for ACL packets, and the buffer sizes \((B_S\) and \(B_M\)) are 15,000 bytes. The use of buffers with a finite size is necessary to perform steady-state simulations in overload conditions.

In Figure 7.5 we plot the aggregate throughput that is achievable in the symmetric and asymmetric cases. It is known that the round robin polling algorithm is the best policy to use when the system is symmetric and completely loaded, and the plotted curves confirm that. However, it is also clear that the round robin polling algorithm is very inefficient under asymmetric conditions because the master continuously polls slaves that have no traffic to send, and this behavior implies bandwidth wastage. In the asymmetric scenario, the Slave 1 local queue saturates, i.e., there are packet losses due to buffer overflow, when the offered load is equal to 400 kbps. By increasing the offered load beyond 400 kbps, the throughput performance increases very slowly.

These results point out the ineffectiveness of round robin scheduling in meeting the requirements of a WPAN highly dynamic scenario. The definition of an efficient scheduling algorithm for Bluetooth is an open research issue. This issue is discussed in [8, 9, 23].

### 7.3 TECHNOLOGIES FOR HIGH-SPEED WLANs

In the past few years, the use of wireless technologies in the LAN environment has become more and more important, and it is easy to foresee that wireless LANs (WLANs) will be the solution for home and office automation. WLANs offer high flexibility and
ease of network installation with respect to wired LAN infrastructures. A WLAN should satisfy the same requirements typical of any LAN, including high capacity, full connectivity among attached stations, and broadcast capability. However, to meet these objectives, WLANs should be designed to face some issues specific to the wireless environment, like security, power consumption, mobility, and bandwidth limitation of the air interface.

Two main standards exist for WLAN: IEEE 802.11 and HiperLAN. HiperLAN (high-performance radio local area network) is a family of standards promoted by the European Telecommunication Standard Institute (ETSI) [15]. The most interesting standard for WLAN is HiperLAN/2. The HiperLAN/2 technology addresses high-speed wireless networks, i.e., those in which data rates range from 6 to 54 Mbit/s. Thus, the technology is suitable for interconnecting portable devices to each other and to broadband core networks such as IP, ATM, and UMTS. Infrastructure-based and ad hoc networking configurations are both supported in HiperLAN/2. HiperLAN/2 is designed to appropriately support data transport characterized by a quality of service (QoS). More details on this technology can be found in [27].

In this chapter, we focus on the IEEE 802.11 technology, as it is mature from an industrial standpoint: IEEE 802.11 cards and access points (both for PC and PDA) are produced by several manufacturers. On the other hand, to the best of our knowledge, HiperLAN is still at the prototype level.

IEEE 802.11 is the standard for wireless local area networks promoted by the Institute of Electrical and Electronics Engineers (IEEE).

The IEEE 802.11 technology operates in the 2.4 GHz industrial, scientific, and medicine (ISM) band and provides wireless connectivity for fixed, portable, and mobile stations within a local area. The IEEE 802.11 technology can be utilized to implement both wireless infrastructure networks and wireless ad hoc networks.

Mandatory support for asynchronous data transfer is specified as well as optional support for distributed time-bounded services, i.e., traffic that is bounded by specified time delays to achieve an acceptable quality of service (QoS).

### 7.3.1 IEEE 802.11 Architecture and Protocols

The IEEE 802.11 standard defines a MAC layer and a physical layer for WLANs (see Figure 7.6). The MAC layer provides to its users both contention-based and contention-free access control on a variety of physical layers. The standard provides two physical layer specifications for radio (frequency hopping spread spectrum, direct sequence spread spectrum), operating in the 2400–2483.5 MHz band (depending on local regulations), and one for infrared. The physical layer provides the basic rates of 1 Mbit/s and 2 Mbit/s. Two projects are currently ongoing to develop higher-speed PHY extensions to 802.11 operating in the 2.4 GHz band (Project 802.11b, handled by TGb) and in the 5 GHz band (Project 802.11a, handled by TGa); see [19].

The basic access method in the IEEE 802.11 MAC protocol is the distributed coordination function (DCF), which is a carrier sense multiple access with collision avoidance (CSMA/CA) MAC protocol. Besides the DCF, the IEEE 802.11 also incorporates an optional/additional access method known as the point coordination function (PCF). PCF is an access method similar to a polling system and uses a point coordinator to determine
which station has the right to transmit. The basic access mechanism is designed to support best effort traffic, like Internet data, that does not require any service guarantees. In scenarios in which service guarantees are also required, the PCF access method must be used. Below, we first describe the DCF access method, and then we present the PCF extension.

**IEEE 802.11 DCF**

The DCF access method, hereafter referred to as “basic access,” is summarized in Figure 7.7. When using the DCF, before a station initiates a transmission, it senses the channel to determine whether another station is transmitting. If the medium is found to be idle for an interval that exceeds the distributed interframe space (DIFS), the station continues with its transmission.* On the other hand (when the medium is busy), the transmission is deferred until the end of the ongoing transmission. A random interval, henceforth referred to as the “backoff interval,” is then selected, which is used to initialize the “backoff timer.” The backoff timer is decreased for as long as the channel is sensed to be idle, stopped when a transmission is detected on the channel, and reactivated when the channel is sensed to be idle again for more than a DIFS. The station transmits when the backoff timer reaches zero.

The DCF adopts a slotted binary exponential backoff technique. In particular, the time immediately following an idle DIFS is slotted, and a station is allowed to transmit only at the beginning of each slot time, which is equal to the time needed at any station to detect the transmission of a packet from any other station. The backoff time is uniformly chosen in the interval (0, $CW$) defined as the “backoff window,” also referred to as the “contention window.” At the first transmission attempt, $CW = CW_{\text{min}}$, and it is doubled at each retransmission up to $CW_{\text{max}}$. In the standard [21] the $CW_{\text{min}}$ and $CW_{\text{max}}$ values depend on the physical layer adopted. For example, for frequency hopping, $CW_{\text{min}}$ and $CW_{\text{max}}$ are 16.

*To guarantee fair access to the shared medium, a station that has just transmitted a packet and has another packet ready for transmission must perform the backoff procedure before initiating the second transmission.
and 1024, respectively (note that CSMA/CA does not rely on the capability of the stations to detect a collision by hearing their own transmission). Immediate positive acknowledgements are employed to ascertain the successful reception of each packet transmission. This is accomplished by the receiver (immediately following the reception of the data frame), which initiates the transmission of an acknowledgment (ACK) frame after a time interval, the short interframe space (SIFS), which is less than the DIFS. If an acknowledgment is not received, the data frame is presumed to have been lost and a retransmission is scheduled. The ACK is not transmitted if the received packet is corrupted. A cyclic redundancy check (CRC) algorithm is adopted to discover transmission errors.

After an erroneous frame is detected (due to collisions or transmission errors), the channel must remain idle for at least an extended interframe space (EIFS) interval before the stations reactivate the backoff algorithm.

The MAC layer also defines virtual carrier sensing: the messages convey the amount of time the channel will be utilized to complete the successful transmission of the data. This information is used by each station to adjust a network allocation vector (NAV) containing the period of time the channel will remain busy.

The basic access mechanism can be extended by a medium reservation mechanism, also referred to as a floor acquisition mechanism, named “request to send/clear to send” (RTS/CTS). In this case, after gaining access to the medium and before starting the transmission of a data packet itself, a short control packet (RTS) is sent to the receiving station announcing the upcoming transmission. The receiver replies to this with a CTS packet to indicate readiness to receive the data. This mechanism can be used to capture the channel control before the transmission of long packets, thus avoiding “long collisions.” In addition, the RTS/CTS mechanism solves the hidden station problem during the transmission of the user data [21]. Further considerations on the protection provided by the RTS/CTS mechanism against the hidden terminal problem can be found in Chapter 27 of this handbook.

7.3.2 IEEE 802.11 Performance

The physical layer technology determines some network parameter values, e.g., SIFS, DIFS, and backoff slot time. Results presented below are obtained by assuming the fre-
frequency hopping, spread spectrum technology at 2 Mbps transmission rate. Table 7.3 shows the configuration parameter values of the IEEE 802.11 WLAN analyzed below.

### IEEE 802.11 Protocol Capacity

The IEEE 802.11 protocol capacity was extensively investigated in [13]. In the following, the main results of that analysis will be summarized. Specifically, in [13] the theoretical throughput limit for the IEEE 802.11 network is analytically derived (i.e., the maximum throughput that can be achieved by adopting the IEEE 802.11 MAC), and compared with the real protocol capacity. These results show that, depending on the network configuration, the standard protocol can operate very far from its theoretical limits. Specifically, as shown in Figure 7.8, the distance between the IEEE 802.11 and the analytical bound increases with the number of active networks, $M$.

#### Table 7.3: WLAN configuration

<table>
<thead>
<tr>
<th>SIFS</th>
<th>DIFS</th>
<th>Backoff slot time</th>
<th>Bit rate</th>
<th>Propagation delay</th>
<th>Stations</th>
<th>$CW_{\text{min}}$</th>
<th>$CW_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>28 μsec</td>
<td>128 μsec</td>
<td>50 μsec</td>
<td>2 Mbps</td>
<td>1 μsec</td>
<td>10, 50, 100</td>
<td>32</td>
<td>256</td>
</tr>
</tbody>
</table>

![Figure 7.8: IEEE 802.11 protocol capacity.](image)
Results presented in Figure 7.8 show that the performance of IEEE 802.11 is negatively affected by an increase in network congestion. This is a typical behavior of random access algorithms that can be partially solved by using stabilization algorithms. These algorithms tune the protocol parameters using feedback from the network. One of the most well-known algorithms of this class is named pseudo-Bayesian [28]. Extending the Rivest approach [28], stabilization algorithms for the IEEE 802.11 have been proposed in [12], [8], and [7]. These works propose different approaches to maintaining the IEEE 802.11 capacity level close to the theoretical bound for all network congestion levels.

The IEEE 802.11 capacity analysis presented above is performed by assuming that the network operates under asymptotic conditions (i.e., each LAN station always has a packet ready for transmission). LANs generally operate under normal conditions, i.e., the network stations generate an aggregate traffic that is lower (or slightly higher) than the maximum traffic the network can support. Under these load conditions, the most meaningful performance figure is the MAC delay, i.e., the time required for a station to successfully transmit the packet at the head of its transmission queue [14]. Results below are obtained by assuming that a station alternates between idle and busy states. State changes occur according to an on/off Markov chain. Specifically, after each successful transmission, a station remains in the on state (i.e., busy state) with probability 0.9. At the end of a transmission, a station in the off state (i.e., idle state) changes its state to on with probability \(x\). By increasing the average sojourn time in the off state, we model a decrease in the network load.

Two sets of experiments were performed corresponding to a traffic generated by 50 stations, made up of short (2 slots) and long (100 slots) messages, respectively. Figure 7.9 (which plots the average MAC delay versus the channel utilization) highlights that, for light load conditions, the IEEE 802.11 exhibits very low MAC delays. However, as the offered load approaches the capacity of the protocol, the MAC delay sharply increases and becomes unbounded. This behavior is due to the CSMA/CA protocol. Under light-load conditions, the protocol introduces almost no overhead (a station can immediately transmit as soon as it has a packet ready for transmission). On the other hand, when the load increases, the collision probability increases as well, and most of the time a transmission results in a collision. Several transmission attempts are necessary before a station is able to transmit a packet, and delays tend to be unbounded.

For this reason, the IEEE 802.11 DCF mode is not suitable for the transmission of delay-sensitive data. Support for users that require quality of service guarantees is provided in IEEE 802.11 by the PCF operational mode. PCF is designed to coexist with the DCF. Best effort users still transmit data using the DCF mode, whereas QoS-sensitive users exploit the PCF mode.

**Point Coordination Function**

The point coordination function guarantees frame transmissions in a contention-free way. This functionality, as shown in Figure 7.6, must be implemented on top of the DCF, and can be used only in infrastructure networks. To determine which station can transmit, the PCF uses a point coordinator (PC) that it is usually implemented in the access point. Stations that use PCF (stations CF_Aware) are recorded in a list managed by the point coordi-
nator, which, through polling techniques, guarantees to these stations a contention-free access to the channel.

In an IEEE 802.11 WLAN, periods under the DCF functionality can be interleaved to periods in which the control is entrusted to the PCF modality. The frequency with which these periods alternate is specified by the CFP_Rate parameter, whose value is selected by the point coordinator. Every contention-free period begins with the transmission of a beacon frame (by the point coordinator), whose main target is the stations’ synchronization. The beacon transmission has higher priority than data transmission. This is obtained by adopting for the beacon transmission an interframe space, named PIFS, that is shorter than the DIFS.

The point coordinator, with the beacon frame, transmits the estimated length of the contention-free period, CFP_Max_Duration, which is used by the stations in the cell to disable the DCF transmissions. This is achieved by setting the NAV to the CFP_Max_Duration value. At the end of the contention-free period, the point coordinator sends a special message that clears the NAV in all stations; hence, it implies switching to the DCF modality.

During the contention-free period, the point coordinator sends polling messages to the CF_Aware stations, enabling contention-free transmission. Each polled station after a SIFS can access the medium and it may choose to transmit a frame to the PC or to another

Figure 7.9  IEEE 802.11 MAC delay.
station in the cell. If the polled station does not reply, the point coordinator, after a PIFS, issues a poll to the next station to be polled according to its scheduling algorithm. The scheduler in the point coordinator is not defined in the standard and is an open research issue.

As in DCF, all data transmissions are acknowledged by the receiving station. For example, as shown in Figure 7.10, Station 1, after receiving the poll from the point coordinator, waits a SIFS and accesses the medium to transmit data to Station 2. Station 2, after receiving the frame, will send, as in DCF mode, the ACK to Station 1. At the end of this station-to-station transmission, Station 1 has to send back to the PC the ACK of the last data message received from the PC. In Figure 7.10, the case of a direct communication between the PC and Station 3 is shown.

To prevent that the point coordinator lockout all DCF traffic (by repeatedly issuing polls), IEEE 802.11 defines a superframe structure. During the first part of this interval, the point coordinator may issue polls, then it idles for the remaining part of the superframe, allowing DCF-based transmissions.

7.4 THIRD-GENERATION CELLULAR SYSTEMS: UMTS

The growing demand for cellular networks providing both high-rate data services and better spectrum efficiency is the main driver for the deployment of third-generation mobile radio networks, often called “3G.” Considerable efforts toward 3G network standardization have been carried out simultaneously in Europe and the United States. Two main standards have been developed. In Europe, the ETSI is developing the UMTS standard for 3G systems and the ITU is developing the IMT-2000 standard for 3G systems, with small differences in the specification of radio interface.

The main objectives for the IMT-2000/UMTS systems are [32]:

![Figure 7.10 Data transmission and related acknowledgment between stations operating in the PCF mode.](image-url)
- Full coverage and mobility for rates up to 144 Kb/s. Up to 2 Mb/s rates for low mobility and limited coverage.
- Use of different sized cells (macro, micro, and pico) for indoor and outdoor applications, with seamless handover between them.
- High spectrum efficiency compared to existing systems.

To achieve these targets, extensive investigations have identified the code division multiple access (CDMA) as the multiple access scheme for the 3G air interface.

CDMA assigns to each user a unique code sequence that is used to code data before transmission. If a receiver knows the code sequence related to a user, it is able to decode the received data. Several users can simultaneously transmit on the same frequency channel by adopting different code sequences. A low cross-correlation among these codes is the main requirement for enabling a receiver to successfully decode the user-generated data. Codes with zero cross-correlation are referred to as orthogonal codes. In particular, the most promising CDMA technique for third-generation systems is the direct sequence (DS)-CDMA, in which the original digital data to be transmitted at time $t$ on a sender–receiver connection, say $b(t)$, is multiplied by a wide-band digital code $c(t)$, which represents the code sequence to be used for that particular connection at time $t$ (see Figure 7.11).

The number of code bits used to represent a single information bit, i.e., the ratio between the chip rate [the bit rate of $c(t)$] and the bit rate of $b(t)$, is known as spreading factor (SF).

![Figure 7.11 Basic spread spectrum.](image)
The different standardization groups have proposed two multiple access schemes based on the DS-CDMA principle for the air interface: the wideband CDMA (W-CDMA) and the time-division CDMA (TD-CDMA). The W-CDMA adopts a DS-CDMA in which all users transmit in the same frequency channel. In W-CDMA systems, the SF can be very large (up to 512), and this is the reason why this technique is called wideband. The TD-CDMA is based on a hybrid access scheme in which each frequency channel is structured in frame and time slots. Within each time slot, more channels can be allocated and separated from each other by means of the DS-CDMA. The number of codes in a time slot is not fixed but depends on the rate and SF of each physical channel.

The multiplexing of the downlink and uplink traffic is implemented with different mechanisms in the W-CDMA and in TD-CDMA. The W-CDMA implements a frequency division duplexing (FDD) mode, in which the uplink and downlink traffic are separated in frequency by using different bands. In this case, a physical channel is identified by a code and one frequency. The TD-CDMA implements a time division duplexing (TDD) mode, in which the uplink and downlink traffic are separated by different time slots assigned to them within the frame.

Figure 7.12 shows the frequencies allocated in Europe for UMTS. The figure clearly indicates that there are two paired 60 MHz bands reserved for W-CDMA. A 35 MHz band, subdivided in two unpaired bands, is reserved for TD-CMDA. Finally, the figure also shows bands reserved for satellite services (SAT).

In the rest of this chapter, we focus on the TD-CDMA technique defined by the UMTS standard. This technique is the most suitable and flexible mechanism for the integration of Internet services in future 3G systems. In fact, the frame structure is tailored to the asymmetry of uplink and downlink Internet traffic, with respect to the FDD, in which the same bandwidth is assigned to the downlink and uplink directions.

### 7.4.1 UMTS Terrestrial Radio Access Network

Three main segments constitute the UMTS system: the core network (CN), the UMTS terrestrial radio access network (UTRAN), and the user equipment (UE), as shown in Figure 7.13.

![Figure 7.12](image-url)  
UMTS frequencies.
The CN is the fixed infrastructure, also called backbone, and it is responsible for the set-up and the control of communication, and the overall management of the users, from mobility to billing. Moreover, the CN is designed to be radio-technology-independent, in the sense that it remains unchanged, whatever the radio access technique.

The UTRAN accomplishes all the radio-dependent functions characterizing the UMTS system, and it represents the radio interface between the mobile user and the network. In particular, UTRAN is constituted by a set of radio network controllers (RNCs) that are connected to the CN. Each RNC is the controller for a group of adjacent base stations (named Node B in the UMTS terminology), e.g., it accomplishes the procedures for handover decisions and macrodiversity management. The macrodiversity mechanism permits a mobile user to have the same connection active through more than one base station with different codes. This functionality enables soft handover, in which the connection continuity is guaranteed through the path multiplicity between the mobile terminal and the “bridging point” in the network. Another particular functionality of the UMTS system is the interfrequency handover. As stated in Section 7.1, the UMTS system is organized in a hierarchical cell structure, with picocells for indoor environments, microcells for urban environments, and macrocells for rural environments. This structure can guarantee maximum area coverage with different mobile user densities. Each layer of this hierarchy has a different frequency assigned to avoid interlayer interference. Hence, the UMTS system has to locate the user in the appropriate cell, according to its mobility pattern, and eventually manage the change of level hierarchy when the user’s mobility pattern varies.

The base station is responsible for the management of radio resources within a single cell, and it generates the radio signal delivered to each user located in its coverage area.

In the following subsection, we concentrate on the logical structure of radio interface
between mobile users and base stations with the aim of introducing the mechanisms and the services that the MAC layer provides to realize the radio resource allocation.

The Radio Interface

The radio interface specifies how both user data and signaling information has to be exchanged between the mobile user and the network [1]. Depending on the user–data transfer requirements, three different operational modes have been defined:

1. Transparent data transfer—provides only segmentation and reassembly procedures
2. Unacknowledged data transfer—provides error and duplicate detection, but no attempts at recovering corrupted messages
3. Acknowledged data transfer—provides a guaranteed delivery of messages from/to upper layers

Below, we will focus on the radio interface adopted in the TD-CDMA system. This interface is named UMTS terrestrial radio access—time division duplex (UTRA-TDD). To support data connections with different bit rates, the TD-CDMA system utilizes coding sequences with variable spreading factor (SF). Since the chip rate is fixed at 3.84 Mchip/s, the lower the SF value, the higher the data rate. In TD-CDMA, the spreading factor may take values 1, 2, 4, 8, and 16. To limit the interference between users transmitting in the same time slot, we need to assign them orthogonal codes, i.e., with zero cross-correlation. This is obtained in UMTS by adopting orthogonal variable spreading factor (OVSF) codes [3, 4]. By using codes with SF = 16, we can have, in principle, up to 16 simultaneous transmissions in the same time slot. Each transmission can deliver a unit quota of traffic, referred to as a resource unit (RU). In general, with SF = x we can have (in the same time slot) only x simultaneous transmissions, which can deliver data corresponding to up to 16/x RUs. To simplify the presentation, hereafter we will always assume SF = 16. Whenever, we wish to assign to a connection a SF less than 16, say x, from a logical standpoint, it is equivalent to assigning to that connection x orthogonal codes with SF = 16.

In the TD-CDMA access mode, the channel is structured in frames of 10 msec length, and each frame contains 15 time slots. Therefore, the radio channel can be modeled by the time-code matrix shown in Figure 7.14. In this matrix, the (i, j) element corresponds to the RU associated to the i-th code (with SF = 16) in the j-th time slot of the frame.

The RUs of the matrix are used to transmit signaling traffic and users’ data both in the uplink and downlink directions. UTRA-TDD has several signaling channels [2, 5], but for the purpose of the following discussion we only introduce the following:

- Synchronization Channel (SCH): downlink channel that delivers all the necessary information to guarantee the synchronism between mobile users and the base station. This synchronism is fundamental to correctly interpreting the frame structure and to identifying the location of the broadcast channel within the frame.
- Broadcast Channel (BCH): downlink channel that delivers specific information related to the cell, such as the frame structure and random access channel location within the frame.
Random Access Channel (RACH): uplink channel that delivers the users’ requests for a new connection. This channel is shared among all the users located in the cell; hence, an appropriate contention control mechanism is needed.

Forward Access Channel (FACH): downlink channel that delivers control information to a mobile user. The control information is mainly related to the RUs assigned (by the scheduler in the base station) to that user for data transmission.

As it is clear from the channel description, only the SCH channel must have a fixed position inside the frame. For the purposes of our discussion on TD-CDMA, we will assume that signaling channels will have a fixed position inside the matrix corresponding to the first and last two slots of each frame (see Figure 7.14). The assignment of the remaining RUs to transmit the user traffic is managed by the scheduling algorithm that is implemented in the base station.

In the following, we will consider two traffic classes: voice and Internet data. The former requires a connection-oriented service with low delay (20 msec) and low bit error rate (BER) requirements (10E-3). Internet data is well delivered by a connectionless service with no stringent delay requirements.

Each time a new voice connection must be activated, the voice terminal issues a request to the base station using the shared RACH channel. Once the request is received by the base station and the UMTS network has accepted the new connection, the resources allocated to the new connection on the UTRA-TDD access channel remain reserved until the connection is closed. On the other hand, the Internet traffic is highly bursty and to manage the channel in an efficient way, resources must be assigned to an Internet user only when it has data to transmit. Several allocation policies can be adopted for Internet traffic. In the following, we follow the approach proposed in [6]. According to this approach, RUs are allocated to the Internet traffic on a frame-by-frame basis. In each frame, RUs not reserved for voice traffic are fairly assigned to active data users (users with data waiting to be transmitted), i.e., each active user receives the same number of RUs. Therefore, Internet users access the RACH to request RU allocation only at the beginning of each busy period of their trans-
mission buffer [25]. When this request is received, the base station will assign RUs to this user in all subsequent frames unless it receives an empty buffer indication.

According to our assumptions, the first time slot in each matrix is reserved for the RACH and the users transmit on this channel randomly, choosing one of the 16 possible orthogonal (SF = 16) codes. Hence, collisions may occur. An Aloha-like algorithm is used to manage this shared channel [6].

**Scheduling Algorithm for UTRA-TDD**

Once the RU allocation requests have been received by the base station, the scheduling algorithm assigns the available resources among the requesting users. The first big problem that a scheduler has to cope with (in a TD-CDMA environment) is soft degradation. Soft degradation makes the number of RUs that can be assigned in a time slot lower than the number of available codes, i.e., 16 with our assumption of SF = 16. This is a characteristic of CDMA systems. In these systems, there is no hard limitation of resources but a soft degradation of transmission quality. A new connection can always be set up if there is an available code because there is no a priori fixed capacity limit. The only bound is provided by the quality criteria adopted by the system, which assesses the maximum interference level acceptable. That is, the simultaneous transmission of different signals over the same bandwidth can take place as long as the interference level does not exceed a given threshold. It is worth noting that in a real environment, it is possible to observe interference between users at the receiving side, even if orthogonal codes are assigned to the users.

Therefore, an UTRA-TDD scheduling algorithm must be both traffic- and interference-adaptable. The latter requirement is tightly connected with soft degradation. The target of the scheduling algorithm is to be fair and efficient. “Fair” means equal sharing of the transmission resources among users of the same class. “Efficient” means that it must optimize some performance figure. For example, for Internet traffic we have chosen to maximize the throughput per frame, i.e., the amount of data successfully transmitted on the radio channel in a frame period.

The most challenging part of the scheduling algorithm is that related to Internet data traffic. For this reason, in the following we assume that we have only Internet data traffic to transmit on the radio channel. In the general case in which voice traffic (and other high-priority traffic) is present, the RUs assigned to the Internet data are those not reserved for the high-priority traffic.

The scheduling algorithm proposed and analyzed in [6] is based on the three steps shown in Figure 7.15. In the first step of the algorithm, the time slots available for Internet data are partitioned in two subsets, UL and DL, for transmission in the uplink and downlink directions, respectively. This partition is created according to a fairness criterion by taking into consideration the number of active users (or, equivalently, nonempty transmission buffers) in the uplink and downlink directions.

Once this partition is generated, for both the uplink and downlink directions, the second step of the algorithm computes (independently for each direction) the number of RUs per time slot (see the second step of Figure 7.15) that can be used for data transmission, taking into consideration the soft degradation phenomenon.

For ease of presentation, let us assume that we know the characteristics of the system interference when \( k \) (\( k \leq SF \), assuming \( SF = 16 \)) codes are used. Specifically, let us indi-
cate with $Pe(k)$ the probability that a RU is discarded at the receiver due to transmission errors when $k$ codes are used in the same slot.

From the $Pe(k)$ knowledge, the algorithm selects the optimal value of $k$, say $k^*$, that maximizes the throughput (see the second step of Figure 7.15) according to the following formula:

$$T(k^*) = \max\{T(i), i = 1, 2, 4, 8, 16\}$$

where $T(i) = E(\text{RUs correctly delivered when } i \text{ codes are used}) = [1 - Pe(i)] \times i$.

Finally, the third step of the scheduling algorithm assigns the same number of RUs to all active users.

Note that hereafter we do not discuss the cases in which the number of RUs per user is not an integer number. The algorithm can be easily extended to manage these cases.

Unfortunately, the assumption that we know a priori the characteristics of the system interference is very far from the real cases. In a CDMA system, the overall system interference depends on a multiplicity of factors:

- The users’ mobility pattern, since the system shows a different behavior depending on the user speed
- The coverage area of each base station
- The intercell interference due to a reuse factor equal to one. It means that each base station utilizes all the radio resources of the network, and it could happen that two adjacent base stations have users that either transmit with the same code, or receive on the same code. To minimize this kind of interference, each station uses a different scrambling sequence to make the signals coming from different cells orthogonal [6]
- The intracell interference due to the simultaneous use over the same radio channel of a variable number of codes, depending on the rate and SF of each physical channel.

Therefore, it is not possible to analytically characterize the system interference; at best we can try to estimate the $Pe(k)$ function by observing the number of RUs correctly received. The estimation of the $Pe(k)$ function is extensively studied and evaluated in [6].
Two main classes of estimation algorithms have been proposed: worst-case and moving-average estimators. The former are very simple, as they use the $Pe(k)$ value related to the last frame. The latter are real estimators, as all the system history contributes to the current $Pe(k)$ estimates.

For example, Figure 7.16 shows the system throughput by assuming that the system interference characteristics generate linear $Pe(k)$ values, i.e., $Pe(0) = 0$, $Pe(16) = 1$, and that the function increases linearly among these two extremes. In the figure, we also show the ideal case results that have been analytically derived by using the $Pe(k)$ formula. As shown in the figure, it is possible to design $Pe(k)$ estimators that drive the system close to its ideal behavior, and that also work well with a channel that shows a rapid change of the interference levels.

ACKNOWLEDGMENTS

The authors wish to thank Graziano Bini for his valuable comments on the manuscript.

REFERENCES

15. ETSI Technical Report 101 683, V1.1.1, Broadband radio access networks (BRAN): High performance local area network (HiperLAN) Type 2; System Overview.


