1

Introduction to Mobile-Phone Systems

Our background in mobile-phone systems has led us to the conclusion that an understanding of such systems helps developers to exploit to the full opportunities for their applications. This chapter will give an overview of such systems.

1.1 Wireless Technologies

Before starting, we must consider the fact that although a wide variety of wireless technologies exist, not all of these are suitable for use by someone who is moving around. Many wireless systems are wrongly described as mobile (and not just by the general public). We must therefore ask the question: what defines a system as mobile? The International Telecommunications Union (ITU) and the European Telecommunications Standards Institute (ETSI) have a very specific definition: ‘The term mobile can be distinguished as applying to those systems designed to support terminals that are in motion when being used.’

This definition precludes such systems as the Wireless Local Area Network (WLAN) technology 80211.b which is not, at present, capable of supporting ‘terminals’ in motion; it is classed as a portable, or nomadic, system, despite Intel’s publicity for its Centrino technology. At present, the only systems that can claim to be truly mobile are the mobile telecommunications networks, irrespective of whether they are using first, second or third generation (1G, 2G or 3G) technology.

As we will see in this chapter, not only does the fact that a mobile phone can be used when in motion affect the definition of the system
of which it forms a part but also the degree of motion greatly affects data transfer rates available for a given application. The specification for 3G systems, for instance, incorporates a steadily decreasing data-rate capability (expressed in bits per second (bps)) as the speed at which the mobile phone is traveling increases:

- A phone moving at 0 km/hr can achieve data rates of up to 2 Mbps.
- A phone moving at 3 km/hr, data rates up to 384 Kbps.
- A phone moving at 30 km/hr, data rates up to 144 Kbps.
- A phone moving at 150 km/hr, data rates up to 64 Kbps.

### 1.2 Cellular Systems

Modern mobile-phone systems are cellular systems, and mobile phones are referred to as cellular phones in some parts of the world. A cellular system divides geographic regions into cells. A cell is the area of coverage within which a transceiver provides a wireless link to mobile phones. Each cell is assigned a unique identification number which identifies its geographical location. Cellular systems were developed in response to the limited availability of radio spectrum. Because only a finite part of the radio spectrum is allocated to mobile-phone systems, only a finite number of users can be supported. To overcome this limitation, cellular systems effectively re-use the available frequency spectrum from one cell to another and so accommodate many more users than would otherwise be possible over the required area of coverage.

The concept of frequency re-use, as employed by a cellular network, introduces the possibility of interference between cells that use the same portion of the frequency spectrum. Mobile-phone networks minimize this possibility by implementing re-use patterns (see Figure 1.1, which shows a re-use factor of seven). Cell sizes and re-use patterns are varied to accommodate the fact that there are more users in some areas than in others. For example, the system may have smaller patterns in urban areas with a large number of potential users and larger patterns in rural areas with smaller numbers of users. For simplicity, cells are often depicted as hexagons but in reality they are often very irregular formations that are dependent on the terrain which in turn affects propagation (see Section 1.6).

As a mobile-phone user moves from one cell to another, the process of maintaining the connection is known as **handover**. The mobile-phone
system monitors the quality of the signal between a given mobile phone and a cell controller or base station against the quality of the signal that would be available if the mobile phone were connected to a base station in an adjacent cell. If the quality of the connection falls below a pre-determined threshold, the connection is handed over to the new cell. This handover process results in a momentary loss of connection. This loss was not a problem for voice-orientated first and second generation mobile-phone systems. The 3G mobile-phone system is primarily data-orientated and a system that allows a loss in connection could result in the failure of transfer of large data blocks. For real-time systems, this problem has been solved in intra-system handover (between cells of the same system, for example from 3G cell to 3G cell) using a ‘make before break methodology’. It remains an issue for inter-system handover (between cells of different systems, for example a 2G cell to a 3G cell); streaming services, such as video calls, cannot be transferred in this situation.

The portion of the frequency spectrum allocated to a cell must be further divided, both to support multiple users within a cell (the multiple-access method) and to separate the connections to and from each mobile phone (the duplex method). In other words, we must keep separate the
calls from different users in the cell and, within each call, we must separate the user’s voice from that of the person they are talking to. In Section 1.4, we discuss multiple-access methods, which follow the changes in mobile-phone-system generations, but first we consider the basic components of mobile-phone systems.

### 1.3 Elements of a Mobile-Phone System

Figure 1.2 shows the basic components of a mobile-phone system and their purpose. The base station (the antenna and associated electronics that control a cell) provides a wireless connection to mobile phones in a given cell and consists of a transceiver and its associated transmitting equipment. Transceivers can be omni-directional (able to transmit in all directions around the transceiver) or sectored (directed at a specific area). Sectored transceivers are able to provide targeted coverage to areas where there are likely to be considerable numbers of users, for example, a football stadium. Providing good coverage represents a considerable monetary investment by the mobile-phone-system operator and is a means by which one operator can differentiate their service from that of a competitor. The base-station controller manages a number of cells and is responsible for handover between cells. The operational center is where the mobile-phone-system operator manages the network. It is responsible, amongst other things, for authentication, billing, maintaining information on the current location of its active users, and providing the interface between other mobile networks and the wired telephone network.

![Figure 1.2 Components of a mobile-phone system](image)
When a user makes a call, a channel is established with the base station, which relays the information to the control center. The control center directs the call to its required location. If the call is to another mobile phone, the control center passes the call to the control center and base station of the cell where the other mobile phone resides. The question often asked is, ‘how does the system know where that mobile phone is located?’ When a user switches on their mobile phone, it registers its location using the identification number of the cell that it is currently in; this information is stored within a database in the operational center. This mechanism allows a mobile phone to be located and connected to calls wherever it may be within a network.

1.4 Keeping Users’ Calls Separate

For the majority of data communications, there is a requirement for several users to simultaneously share a common channel. This channel could be a high-speed optical-fiber link between two countries or the frequency spectrum in a mobile telephone system. For this sharing to occur, we require an access protocol that defines when or how the sharing is to take place. This sharing process is referred to as multiplexing in wired communication systems and multiple access in wireless communications. Three classes of multiple access are particularly relevant to mobile-phone systems; they assign individual users different frequency or time slots, or identify individual users by different codes, as shown in Figure 1.3.

Frequency-division multiple access (FDMA) splits the available spectrum into smaller portions, each of which is then allocated to an individual user. FDMA was used in many of the first-generation (1G) analog mobile phone systems and extended a technique used to transmit multiple calls down a single wire of a telephone network. FDMA is particularly susceptible to frequency-selective fading (a phenomenon that is discussed in Section 1.5), which is why it has been superseded by other techniques.

Time-division multiple access (TDMA) allows a single user to occupy all of the available bandwidth, but only for an allotted period of time. TDMA has improved immunity to frequency-selective fading and was
Figure 1.3  Multiple-access techniques in a mobile-phone system
chosen for some of the second-generation (2G) mobile-phone systems, most notably Global System for Mobile (GSM) communications. Code-division multiple access (CDMA) allows users to share the time and frequency domains simultaneously by allocating each user a unique identification code. This code means that the system is able to separate one call from the other even if two calls are transmitted at the same time. CDMA is the main technique chosen for third-generation (3G) mobile-phone systems as it allows a higher data rate for each user.

Having separated the individual users within a cell, we must also separate the outgoing and incoming traffic (the *uplink* and *downlink*) to each user (for example, for a voice call we separate what the user hears from what the user says). 3G systems use frequency-division duplex (FDD) in which separate frequency bands are used for the uplink and downlink. An alternative is to use time-division duplex (TDD), in which the uplink and downlink transmit on the same frequency but share the resources in time. It is worth noting that TDD is being considered for use in indoor environments, such as museums and shopping centers.

### 1.5 Multipath Propagation

The mobile channel is one of the most challenging environments in which communications systems must operate. This is predominantly due to the fact that mobile phones operate in the microwave frequency range, where the resulting wavelength is in the range of centimeters. This means that any object of that size or greater can obstruct or reflect the signal. Reflections from objects can result in multiple copies of the same signal arriving at a location via different paths. These copies are combined in the channel to form the overall received signal. The lengths of these paths are likely to be different, causing some reflections to be delayed in arriving and resulting in positive or negative additions of the overall received signal – in other words, the signal gets either stronger or weaker because of the reflections. A simple way to view this effect is to consider the signal as a binary sequence of $-1$ and $+1$. When the reflections occur, sometimes two values add together to enhance the signal, giving a result of $-2$ or $+2$; in other cases, $+1$ will be added to $-1$ to give a resultant of 0 (see Figure 1.4). This effect is known as multipath propagation and it can be observed by the user as a change in the signal strength display on their mobile phone as they change position. Negative additions are usually referred to as fades.
Obviously, the faster the speed at which the mobile phone is moving, the greater the change in the quality of the signal. This is one of the reasons that data rates are inversely proportional to the speed at which a mobile phone is moving. However, the positive aspect of multipath propagation is that it enables us to receive a signal on a mobile phone within the heart of a crowded city.
Multipath propagation is frequency-dependent (generally referred to as frequency-selective). Different frequencies experience different effects from path lengths and some frequencies are negatively affected, while others are positively affected for a mobile in exactly the same position – which means two mobile users can experience different system performance even if standing in the same spot. This brings us back to the point we made earlier about FDMA. As users are assigned a small portion of the frequency, they may be completely within a fade area. Using TDMA, which operates over the entire bandwidth, only a small part of the user’s signal would be in fade and the problem can be overcome by the application of error-correcting codes.

1.6 2G Mobile-Phone Systems

The evolution of mobile-phone systems from 1G to 2G is principally viewed as a move from the original analog systems of the late 1970s and early 1980s, in which the transmitted signals consisted of modulated versions of the original speech, to the digital systems of the 1990s, where the speech was converted to a series of ones and zeros by passing the voice signal through a transformation process. There was a plethora of analog standards around the world, amongst them the American Mobile Phone Standard (AMPS) in the USA and the Total Access Communications System (TACS) in the UK, all of which suffered from a number of limitations. The number of users was limited because analog signals cannot be compressed; security was virtually non-existent as analog signals cannot be encrypted and anyone with a cheap scanner could listen in on the phone calls; and error-correction techniques could not be applied as an analog signal has an infinite number of possible levels which cannot be predicted – all of which resulted in higher power requirements from the mobile phone.

In Europe, there were nine different analog standards and no facility for mobile-phone users to ‘roam’ around the continent. Europeans therefore saw the need to adopt a single standard for the entire continent and took the rather brave decision to abandon their original 1G provision for an entirely new system. National phone operators, manufacturers, and regulators came together to produce the standard known as GSM. In the USA, where there was no problem of competing standards, a legacy system was chosen. Digital AMPS (D-AMPS) mobile phones could switch between digital and analog operation. In the 1990s, another 2G system
appeared in the USA, based on CDMA technology; it is now known as cdmaOne. This system proved important in the provision of 3G systems (see Section 1.8).

Network operators started thinking about possible 3G systems around the same time the Internet expansion started and this very much shaped the vision of 3G systems as principally orientated towards data rather than voice. The problem with sending data over 2G systems came from the fact that they used circuit switching to make connections between locations. In a circuit-switched system, the connection between the source and the destination is maintained throughout the duration of a call. This technique is very effective for voice calls where the transmission is fairly constant throughout the call but for data sessions, such as web browsing, where there are large periods of inactivity, it is extremely wasteful of capacity because the connection cannot be utilized by other users. Packet switching, on the other hand, splits data into small packets which can be routed down any available path. This maximizes system utilization, allowing multiple connections to transmit packets simultaneously across the system, as opposed to a single connection blocking the system to other connections (see Figure 1.5).

The first stage of the evolution to 3G systems was the introduction of packet data into the existing 2G network via the General Packet Radio Service (GPRS). This extension of the 2G service, but without the rapid data transfer rates associated with 3G systems, has been called 2.5G systems.

1.7 GPRS Systems

GPRS is an essential part of the migration of GSM services to 3G systems and allows the user to have a constant connection to the Internet.

We have already discussed the need for packet data in Section 1.6. The ability to stay connected is also a requirement that comes from the intermittent nature of the activity of data traffic. In circuit-switched systems, users have to connect to the data service and are charged for as long as they stay connected, regardless of whether they fully utilize the link or not. This leads to a situation where users must go through the long connection and verification procedure every time they wish to access data. In packet-switched data systems, where capacity is allocated as and when it is required, consumers are only charged by the amount of data they transmit and receive. This means the system can allow users
Packet-switching system

Circuit-switching system

Figure 1.5 Circuit-switching and packet-switching systems
to remain logged into the network as long as their mobile phone is switched on.

Comparing Figure 1.6 with Figure 1.2, we can see the addition of two new hardware elements to the existing GSM system. The Serving GPRS Support Node (SGSN) controls the routing of packet data within the mobile network and the Gateway GPRS Support Node (GGSN) connects the mobile network to the general Internet infrastructure. Packet data is therefore added to the existing GSM structure and is separated from voice traffic at the base-station controller, which has a new hardware element, the packet data control unit, for this purpose. All other changes are achieved through software, making GPRS a very cost-effective upgrade because no physical changes were required at the base stations.

Fitting packet data within the GMS structure means that its performance is limited, and the data rates achieved in practice are relatively modest. Voice traffic has priority and data services are provided when spare capacity is available. Furthermore, data rates are generally asymmetric in provision (users tend to receive more data than they transmit) so the uplink can deliver much lower rates than the downlink. Typical values would be in the range of 10 Kbps in the uplink and 40 Kbps in the downlink, depending on phone capabilities. Note that there are no end-to-end delay guarantees for GPRS and latency (that is, the time taken for data to be received from when it was transmitted) could be in seconds.

Application developers wishing to take advantage of GPRS services must carefully consider whether their application will provide effective services when operating in this environment.
1.8 3G Mobile-Phone Systems

While the vision for 2G systems was very much concerned with increasing capacity in terms of the number of users it was able to support, 3G technology was more concerned with increasing the available data rates to those users. Initially, the 3G system was envisaged as a global system. To achieve this, it was proposed to use the 2 GHz frequency band. However, Figure 1.7 shows that there were a number of problems in this band, particularly in the USA, which had already allocated a significant proportion of it for the Personal Communications System (PCS). PCS is the same operational standard as GSM, but operates at higher frequencies (1800 and 1900), hence the need for a tri-band phone when traveling in the USA. Europe and Japan had also allocated part of this band for cordless telephone systems, the Digital European Cordless Telephone (DECT) standard and the Personal Handyphone System (PHS), although these were easily overcome by limiting the allocation of licenses. This meant that any standard using the 2 GHz band was unavailable to the USA and the result was that multiple 3G standards were developed under the auspices of the 3rd Generation Partnership Project (3GPP).

Europe and Japan opted for wideband CDMA (W-CDMA) using frequency-division duplex (FDD) in the two paired bands of the spectrum.

![Figure 1.7 Allocation of 3G frequencies](image-url)
The USA opted for an extension to cdmaOne, which uses multiple bands from this system to produce the so-called multi-carrier CDMA or CDMA2000. Another W-CDMA system is available in the unpaired band of the allocated spectrum. It uses time-division duplex (TDD) and is envisaged for providing 3G services in an indoor environment. Although not truly a 3G system, there is a further extension to GPRS called the enhanced data rates for GSM evolution (EDGE), which modifies the wireless link between a mobile phone and the base station of the GSM/GPRS system to further improve data rates. This standard is also developed by 3GPP.

In Section 1.7, we described GPRS as a stepping stone from 2G to 3G systems. Figure 1.8 shows how W-CDMA is incorporated into the infrastructure shown in Figure 1.6 to provide 3G services.

The two new hardware elements introduced are W-CDMA base stations and the radio-network controller (RNC), which provides much of the same functionality as the base-station controller but with improved handover for data streaming services and provision for the allocation of Quality of Service (QoS). 3G technology can therefore be introduced into the existing system in a controlled manner with the urban areas being serviced first. Indeed it is questionable whether network operators will consider it cost-effective to introduce 3G services into rural areas for some time. This means that network operators must supply dual mode mobile phones (capable of operating in both 2G and 3G modes). This

---

**Figure 1.8** Introduction of W-CDMA into GSM/GPRS systems
also has implications for applications developers if their application is to allow operation over the complete coverage area.

Potential data rates are dependent on a range of factors, such as the number of users in a cell, the noise levels in the cell and the speed of the mobile phone. In previous generations of mobile networks, everyone was effectively provided with the same level of service irrespective of how much they paid. The data-centric orientation of 3G systems allows for the collective effect of service performance to determine the degree of satisfaction of the end-user of the service. Different data services require different levels of support from the system, for example, a video call transmits more data than a voice call and that data must be transmitted at regular intervals to ensure the video appears smooth, whereas web browsing requires intermittent data transfer when the user selects a new page or link.

When connected to a 3G network, a user is assigned a profile which includes essential parameters of the mobile phone’s application to describe its required level of QoS, such as traffic classes, target transfer delay, reliability, guaranteed bit rate and priority. QoS for 3G systems is categorized into four traffic classes as shown in Table 1.1, where the

<table>
<thead>
<tr>
<th>Traffic Class</th>
<th>Characteristics</th>
<th>Typical Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conversational (real time)</td>
<td>Low end-to-end delay</td>
<td>Voice, highly intensive games</td>
</tr>
<tr>
<td></td>
<td>Preserves time relation between packets</td>
<td></td>
</tr>
<tr>
<td>Streaming (real time)</td>
<td>Preserves time relation between packets</td>
<td>Streaming media</td>
</tr>
<tr>
<td>Interactive (non-real time)</td>
<td>Request response pattern</td>
<td>Web browsing</td>
</tr>
<tr>
<td></td>
<td>Preserves data integrity</td>
<td></td>
</tr>
<tr>
<td>Background (non-real time)</td>
<td>Best effort</td>
<td>Background synchronization, downloads, etc.</td>
</tr>
<tr>
<td></td>
<td>Non-time-critical data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Preserves data integrity</td>
<td></td>
</tr>
</tbody>
</table>
main distinguishing factor between these classes is the sensitivity to delay of the expected traffic.

The QoS profiles offered to users are tariff-dependent and represent another variable that can affect the likely take-up of a developer’s application.

## 1.9 IP Multimedia Subsystem

The IP multimedia subsystem (IMS) aims to merge the two successful paradigms of the Internet and cellular systems. IMS is an architecture that defines how IP networks should handle voice and data sessions by replacing circuit-switched telecommunications. It is a service-oriented architecture and employs a distributed component model for the applications running on top of it. This means that it aims to separate the services from the underlying networks that carry them. It originated from 3GPP as a means of allowing 3G mobile-phone-system operators migrating from GSM to deliver more effective data services. Since then it has been adopted by other standards organizations for both wired and wireless networks. With Session Initiation Protocol (SIP) as the backbone of the system it is gaining widespread backing from service providers, mobile-phone vendors, application developers and infrastructure vendors.

The architecture of IMS consists of three layers: the transport layer, the control layer and the application layer. Since IMS can separate the services from the underlying carrier, a GPRS-enabled mobile phone can connect to IMS equally as well as a PC connected via a Digital Subscriber Line (DSL). More importantly in a mobile environment, where a user has the ability to roam, the access-independence of IMS does not only allow both the physical roaming of the user, but also provides the ability for the device to roam between various connection methods. For example, WiFi-enabled mobile phones could switch seamlessly between GPRS and WiFi and users could even switch from using their mobile phone to a PC within the same session and as the same user. Both of these features ensure that the upper layers of the communication system are saved from large amounts of data traffic; in other words, only essential information is passed to the application and all other data is maintained at the physical interface.

In simple terms, IMS can be regarded as turning current mobile-phone systems – with their mixture of circuit- and packet-switched systems – into fully packetized IP networks including all voice traffic. This means that
developers will have greater resources with which to work: IMS is a platform for creating and delivering a wide range of services. Note that it is only responsible for connection, state management, QoS and mobility while the actual data transfer takes place directly between devices.

When IMS infrastructure is available for commercial applications in a few years’ time, applications will be greatly enhanced and may take advantage of SIP, the signaling protocol for IMS, to synchronize, for instance, with a game server so that a multiplayer game can ensure that all players are experiencing the same game action at the same time. Further, a SIP message initiated by the server can send out requests to all mobile clients to take certain actions or wake up from running as a background process and bring updates to the display. This is beneficial for applications in which users are often interrupted.

When all user information is available, IMS can be added to an application without major overheads. The ability to detect the presence of a user can initiate better community features, encouraging users, for instance, to compete against their friends in a particular multiplayer game.

IMS has exciting new services and functions and is a step towards utilizing even more of the potential of mobile-phone systems and services.

1.10 Mobile-Phone Hardware

In the previous sections we focused on the constraints imposed by mobile-phone systems and the environment in which they operate. In this section, we will focus on the resource constraints of mobile phones.

The applications that we will teach you to develop in this book will be designed for mobile phones based on Symbian OS and the S60 user interface. Symbian OS is an open operating system derived from technology initially developed by Psion Software, but subsequently licensed and developed by an affiliation of major mobile-phone manufacturers (more information on Symbian OS and its history can be found at www.symbian.com). Symbian OS supports a range of user interfaces and development languages; in this book we will focus on S60 and the native programming language of Symbian OS, C++.

The S60 user interface was originally designed for one-handed use and, as a result, S60 phones normally offer a standard ITU-T keypad, two soft keys, buttons for starting and ending phone calls (these are not available to application developers), and a directional keypad. There are, however, exceptions to this and Nokia in particular have come up
with a number of non-standard key layouts which could cause problems to the unwary developer. Having limited key options available has a direct bearing on the user interface (UI) design and applications must be developed accordingly. Another limitation to UI design is the small screen size, which means that developers must get their ideas across in a very small area. At present, most S60 phones have screens that are 176 × 208 pixels with 12-bit color resolution, though this is not actually defined within the S60 specification, and future devices may well deviate from it.

The hardware architecture of the S60 mobile phone is currently based around fixed-point digital signal processors (DSP) rather than the more generic floating-point microprocessors commonly found in PCs. This is because mobile phones must principally carry out the algorithmic data manipulation associated with both audio and video coding and are portable devices that must conserve power. Typical processor speeds are in the order of 100 MHz, which both conserves power and prevents the excessive heat generation associated with the high clock speeds of a typical PC processor. Processor speed is a limiting factor for sophisticated graphics applications, particularly those using three-dimensional graphics, and applications of this type will generally drain the phone battery very quickly. While Symbian OS does allow the use of floating-point arithmetic within an application, developers should be aware that it is much less efficient in terms of application speed and power usage than using fixed-point arithmetic.

The memory of S60 mobile phones is generally limited to a few megabytes for applications and about a megabyte of heap memory (the memory available for applications to use while in operation). However, if you are intending to deliver your applications over the air (OTA), you should check the size allowed by operators; on average, it is currently only in the region of 64 KB.

The aim of this book is to give developers an overview of the operation of mobile-phone systems and devices in order to allow them to create novel applications that exploit these systems. Mobile-phone systems are still undergoing rapid technological change and thus we strongly urge developers to keep themselves updated with regard to future trends and directions or they will severely limit their options.