Introduction

After the successful global introduction during the past decade of the second generation (2G) digital mobile communications systems, it seems that the third generation (3G) Universal Mobile Communication System (UMTS) has finally taken off, at least in some regions. The plethora of new services that are expected to be offered by this system requires the development of new paradigms in the way scarce radio resources should be managed. The Quality of Service (QoS) concept, which introduces in a natural way the service differentiation and the possibility of adapting the resource consumption to the specific service requirements, will open the door for the provision of advanced wireless services to the mass market.

Within this context, this chapter introduces the basic framework for the development of the radio resource management strategies, which is the main object of this book. To this end, Section 1.1 analyses the evolution of the mobile communications sector and tries to identify the key socio-economical aspects that could enable a successful deployment of 3G systems. In turn, Section 1.2 provides a description of the basic features of UMTS from the architectural point of view, including the initial architectures of the first releases as well as the evolution towards all-IP networks. Finally, Section 1.3 presents the QoS model that is defined in UMTS, including the identified service classes and the main QoS attributes.

1.1 THE MOBILE COMMUNICATIONS SECTOR

The development of mobile communications has traditionally been viewed as a sequence of successive generations. The first generation of analogue mobile telephony was followed by the second, digital, generation. Then, the third generation was envisaged to enable full multimedia data transmission as well as voice communications. However, the high cost and technical difficulties faced in standardisation and development have led to delays in 3G deployment and, in the meantime, the model of a succession of generations began to break down, first with the intercalation of a 2.5G enabling basic Internet access from mobile terminals, and then with the emergence of public WLAN (Wireless Local Area Network) technologies as potential competitors of the 3G UMTS (Universal Mobile Telecommunications System). In this context, looking at the period 2010–2015, the concept of beyond 3G encompasses a scenario with a variety of interoperating systems, each filling a different niche in the mobile communications market.

Recommendation ITU-R M.1645 defines the framework and overall objectives of future development of IMT-2000 (International Mobile Telecommunications 2000) and systems beyond IMT-2000 for the radio access network. In this respect, the significant technology trends need to be considered. Depending on their development, evolution, expected capabilities and deployment cost, each of these technologies
may or may not have an impact or be used in the future. Moreover, beyond 3G technology is still very immature and a range of alternative scenarios remain possible. As a result, all the forecasts are by definition open to criticism. How mobile communications will evolve over the forthcoming years will depend on the interaction of a number of factors. These include the progress made in developing the various technologies, the emergence of new applications, and the adoption of new services by users. Although the technology is an essential element, a viable business model is clearly a crucial factor.

Information and communication technologies play an important role in determining competitiveness, employment and economic growth. They create new opportunities that at the same time affect existing production, communication and distribution processes. No technological development is possible without an effect upon society. Clearly, no one will deny the evolving nexus between technological innovation and the human condition. Technical devices have never before played such an important role in our daily lives. The development of mobile technologies has been pivotal in this transformation and, consequently, some considerations are discussed in Section 1.1.1. Plausible key factors in future market developments are covered in Section 1.1.2. Furthermore, the complexity of the mobile communications sector is due to a mix of technologies, business models, socio-cultural influences, etc., and therefore we must take notice of market developments in early adopters, such as Japan, described in Section 1.1.3. From this standpoint, the situation and approaches in different regions are covered in Section 1.1.4. The role of technological advances is stressed in Section 1.1.5.

Much analysis covering technical, business and demand-related aspects of what the future mobile communications environment might be like have been produced in different fora. This section collects different perspectives and sources together in order to forecast and/or highlight the key issues in the wireless arena, with the aim of providing a self-contained framework and a broader perspective on the Radio Resource Management problem. In particular, technical reports of the Institute for Prospective Technological Studies (IPTS) of the European Commission [1][2] and ITU background papers [3] and draft reports [4][5] have been considered. The interested reader is directed to these references for more details on these topics.

1.1.1 THE MOBILE EXPERIENCE

The world has witnessed an explosion in the growth of mobile communications in recent years. Year 2002 marked a turning point in the history of telecommunications in that the number of mobile subscribers overtook the number of fixed-line subscribers on a global scale, and mobile became the dominant technology for voice communications.

As a technical device, the mobile phone has become an incredible important part of human life, and a powerful determinant of individual identity. Indeed, the mobile phone has moved beyond being a mere technical device to becoming a key social object present in every aspect of our daily lives. At the same time, the highly personalised nature of the mobile phone has meant that its form and use have become important aspects of the individuality of a phone user. The mobile phone has indeed become one of the most intimate aspects of a user’s personal sphere of objects (e.g. keys, wallet, money, etc.). Both physical and emotional attachment to mobile handsets is increasing. The mobile phone has become somewhat of a status symbol. Mobiles are quickly becoming fashion accessories rather than simply communications devices. The introduction of the mobile phone has also facilitated the balancing of professional and domestic life. In this respect, the mobile phone has become metaphorically an extension of one’s physical self, intrinsically linked to identity and accessibility.

1.1.2 THE BUSINESS CASE

With voice traffic over current GSM (Global System for Mobile communications) and other networks approaching saturation point in many European countries, there is a real opportunity for 3G networks to accommodate the capacity shortage that is likely to emerge in the medium-term. There is as yet a lack of ‘killer applications’ for the mobile Internet in Europe. While MMS (Multimedia Messaging Service) and
adult entertainment have been attractive to consumers, operators may need to realise that simultaneous
efforts must be made to obtain customer preferences from a wide range of demographic, social and
economic backgrounds in order to define market segments of service offerings. A possible weakness,
paradoxically, lies in the cultural and linguistic diversity of Europe, which could work against 3G take-
up. This is because localisation of content could increase the cost of production and subscribers may
have to absorb part of it.

Doubts about the market potential of mobile data and multimedia have lowered expectations for 3G,
and the roll out of 3G services has run into difficulties. As the lack of demand for 3G has shown, it is
extremely difficult to predict the likely market adoption of mobile wireless communications and the
revenues that can be expected. Added to this uncertainty is the potential impact of public WLANs.
However, although operators have been deploying public WLAN networks for some years, most have
been unable to turn them into a profitable business. Some estimations suggest that standalone public
WLAN services will probably not provide a sustainable business in the short-term, despite the free use of
spectrum and the relatively small investments required compared to 3G. The intrinsic problem of
achieving efficient usage of free un-coordinated bandwidth could become critical as more players enter
the field. Nevertheless, WLANs may prove to be of high strategic value and an important source of
competitive differentiation. Even if the direct revenue impact of public WLAN is low, they may be
important for subscriber retention, or as the means by which a fixed line operator could enter the mobile
market.

Viable business models for public WLAN will depend on the cost of access to the backbone network,
security, and charging mechanisms. As a public mobile technology, it could potentially evolve as a
separate competitor to cellular networks in the form of a network of hotspots or it could become more
closely integrated within the cellular network. Although public WLANs cannot substitute entirely for 3G
in terms of functionality, if they are able to offer most of the services users might want from 3G at lower
cost, they may undermine 3G’s business model. Nevertheless, WLANs might stimulate demand for
mobile broadband and create a cohort of users willing to pay to upgrade to higher quality 3G when they
tire of the limited coverage, high demands on battery power, patchwork of hotspot ownership and
congestion of WLAN access points. What seems less likely today, however, in the light of the problems
faced by 3G deployment and in the context of emerging technologies, is a smooth linear transition to a
homogeneous and universal fourth generation (4G) at some point in the medium term.

Considering the length of time that 3G appears to be taking to rollout, it could be overtaken by
alternative technologies such as WLAN, old technologies such as GPRS (General Packet Radio Service),
and increasingly sophisticated pager technology. Licensing problems arising from the multiple patents
held by various parties to the 3G technologies also pose a complex and expensive issue, recalling the
GSM patent problem. Furthermore, since each generation of handheld gadgets contains more and more
complex software, it could turn potential 3G users away because the general consumer is finding it harder
to leverage his knowledge from one gadget to another.

It may also appear that competition between different technologies (in the case of 3G, CDMA2000
versus WCDMA) helps bring down prices. The obvious policy conclusion, therefore, would be to shape
market conditions so as to encourage competition between standards. On the other hand, experiences
from 1G and 2G point to the opposite conclusion. Too much competition between technologies/standards
limits the possibilities of economies of scale, and so the right balance is needed. Similarly, the right
balance is needed to harmonise operators’ and vendors’ diverging strategic visions. However, the fragile
business case suggests efforts should concentrate on creating a dynamic and sophisticated market for
advanced mobile data and voice services based on 3G technologies. If this can be achieved, at the same
time as integrating new technologies to improve the user experience further, the evolutionary path
towards 4G will become clearer and maintain its momentum.

The downturn in the telecommunications sector caused by excessive operator debt and disappointment
over market growth, as well as the extreme cases of vendor financing, makes it highly likely that it will
be more difficult to secure financial backing for new investments in a future generation of mobile
communications systems. It has been suggested that several 3G operators may recoup their investments
slowly, and this will reduce the likelihood of operators investing in 4G by 2011, the date tentatively set by several equipment vendors for its introduction. Instead, for most operators, this investment is likely to be postponed a long way into the future. However, before more accurate predictions of operator investments in 4G can be made, 3G adoption will have to take off. It does not seem likely that a very high-speed mobile data network will gain user acceptance unless successful mobile data applications have been developed and commercialised with 3G.

1.1.3 A LEARNING CASE STUDY: JAPAN

The Japanese market is far more advanced than other regions in terms of the extent of use of cellular mobile data services and terminals. Therefore, it provides one of the few learning experiences that can provide feedback into the design of future mobile communication systems.

In the 2G world, very few countries have been successful with the ‘mobile Internet’. WAP (Wireless Application Protocol) in Europe suffered from low transmission speeds, paucity of content and disenchanted users. Japan, on the other hand, introduced a wide array of mobile Internet services, and witnessed phenomenal growth in usage and subscribers. In fact, Japan made mobile Internet services an integral part of mobile phone ownership, and even made charging for Internet content a reality. The country exhibits the highest total number of mobile Internet users in the world.

NTT DoCoMo launched its Internet connection service, ‘i-mode’, in February 1999. i-mode subscribers can connect to the Internet through special designated handsets. The main services are email, information services and applications such as Internet banking and ticket reservation. Other mobile operators in Japan also began competitive Internet connection services in 1999. In September 2003, there were 78.6 million cellular mobile subscribers in Japan, of which 84% were using some kind of Internet browsing service. In 2003, the average annual revenue per i-mode user was about 200 €, most of which stems from packet transmission charges. The primary use of mobile Internet in Japan is for email: over 83% of mobile subscribers use the mobile Internet for sending and receiving email. Downloading or listening to online music, such as ring tones or tunes, and purchasing online content are other examples of key usages.

Low PC penetration is one of the main factors contributing to the success of mobile networks for Internet access in Japan. Some analysts point to the large number of long-distance commuters using public transport as a stimulus for growth. Nevertheless, a large majority of Japanese use their mobile phone at home to make calls and some surveys also show that the use of the mobile browser in Japan is highest at home (in fact the peak time period for browser usage is after working hours, between 19:00 and 23:00 on weekdays). The introduction of colour display handsets is claimed to be another major driver for the take-up of i-mode services.

Japan has carefully and successfully developed the 2.5G mobile Internet market, thus cultivating the whole innovation system (in terms of usage, operating networks, terminal supply, content development, etc.). This cultivation has not only prepared the Japanese market for 3G services, it has given them first-mover advantages that they can leverage on the international market. Thus, it is expected that market shares of Japanese handset manufacturers and other actors will increase when the transition to 3G (and mobile Internet) takes place elsewhere.

The policies on the introduction of higher-speed 3G services in Japan fixed the number of operators to three per region, due to the shortage of frequencies. The regulator had a total of 60 MHz available for 3G services (uplink and downlink). In order to allocate a minimum of 2 × 20 MHz blocks of spectrum, only 3 licences could be awarded. New as well as incumbent operators were eligible for the licences. Operators were required to cover 50% of the population in the first five years. Only the three incumbent operators, i.e. NTT DoCoMo Group, IDO and Cellular Group (KDDI), and J-Phone Group, applied, and obtained, the three available licences in each region.

NTT DoCoMo was the first operator to launch 3G services in Japan, under the brand name FOMA (Freedom of Mobile Multimedia Access), and based on WCDMA (Wideband CDMA). The full-scale commercial launch of FOMA was initially scheduled for 30 May 2001. Although DoCoMo postponed
the launch until October 2001, it was one of the first operators to launch a 3G commercial service. However, due to the limited service coverage at the time of launch, the fact that the WCDMA system does not have backward compatibility with its 2G service based on the Personal Digital Cellular (PDC) system, relatively short battery life and lack of killer applications (the highly publicised video-phone capability was not a resounding success), it was only by the end of 2002 that 150,000 subscribers were reached. Then, the advent of a flat rate contributed to a very significant increase in the number of subscribers.

High-speed Internet access services based on WLAN were launched in 2002 in Japan. However, it seemed a challenging task to develop a sound business model, attracting a large number of paying users. There are also several WLAN access points offered free of charge by a number of providers. Nonetheless, other types of fixed wireless access services are being launched. A handful of companies are planning to offer a wireless IP (Internet Protocol) phone service for Personal Digital Assistants (PDAs) and WLAN service providers are hoping this will get them out of their current business plan conundrum, but it remains to be seen whether they will be successful or not.

The lack of profitability of WLAN services is likely to persist for some time to come, and for this reason, a number of providers are exploring options to combine or integrate WLAN services with other types of services, notably NTT Communications and NTT DoCoMo. A WLAN service is being offered in combination with its 3G or FOMA service, which typically provides speeds of 384 kb/s so far. Users can benefit from 3G data transmission rates when away from WLAN access points, through the 3G network.

One of the most distinguishing aspects of the Japanese mobile industry is that it is operator-led. Equipment manufacturers and operators work very closely and supply the market with handsets and portable devices in a coordinated effort. The close relationship between manufacturers and operators in Japan accounts in part for the sophistication and availability of handset technology and the take-up of value-added services. Another peculiarity of the Japanese mobile market is the early agreement between content providers and operators. In principle, the mobile operator bills for content, retains a commission, and passes on the majority of the content fees to the content provider.

1.1.4 REGIONAL PERSPECTIVES IN MOBILE EVOLUTION TOWARDS 4G

The European roadmap encompasses a clear tendency towards the development of a future mobile system where heterogeneous technologies, complementing each other in terms of coverage, bit rate and other characteristics, work together in a seamless system to optimise usability for the end user. There is an emphasis on taking advantage of existing and emerging technologies to provide what is, from an end-user perspective, a seamlessly integrated communications environment, with software defined radio as an enabling technology.

Although a European consensus seems to exist on the future diversity of wireless technologies and on the development of services driven by user needs as opposed to technology push, these visions express uncertainty as to the industry structure that will deliver 4G services in the 2010–2015 timeframe, partially motivated by the emergence of new players and the possibility of a fragmented industry. In the short term, 3G in Europe will be driven by mobile operators and especially telecom equipment suppliers.

In Europe, limited experience of advanced mobile data communications is still observed and, for the time being, there are not yet signs of any increase in demand from users for these services (in contrast to Japan, which is the world’s most advanced mobile market). There is clearly a need to abandon the technology push approach that has so far characterised European mobile communications in favour of a more user-focused perspective.

Europe runs the risks of being a late starter in the race to deploy 4G. In this situation, mobile telecommunications equipment will be built cheaply in Asia, causing Europe to fall behind in the production and deployment of mobile communications systems. The development and adoption of 4G in Europe will require the prior large-scale adoption of 3G. While European actors should certainly aim for a leading role in 4G in the future to avoid missing opportunities, efforts should also be made to
consolidate 3G infrastructure as a means of supporting a multitude of coexisting applications and enable the continuous incorporation of emerging standards and technologies. The standardisation made possible by UMTS adoption is an opportunity, but does not mean that other emerging technologies and standards should be ignored. On the contrary, UMTS integration should be the priority in the coming years, encouraging other standards to be made compatible with UMTS, promoting its enhancement and ensuring the removal of any barriers to its adoption. It should include provisions for spectrum regulation harmonisation and interconnection issues, which would allow investments in 3G infrastructure to be recouped without missing the opportunities stemming from technological innovation in other areas.

The US appears to lack a shared industry-wide view of how mobile telecommunications are likely to develop; at the same time, there is no representative body that articulates US visions for 4G. The trend in the US is towards new proprietary technologies deployed over unlicensed spectrum, coexisting with new standards developed for use on both unlicensed and licensed spectrum. At the same time, more unlicensed spectrum is being made available and flexible spectrum management is supporting the interoperability of products and technologies offered by a more fragmented industry. Thus, the US is leading the way in the deployment of potentially disruptive technologies such as public WLAN. The push by some US actors to make further free spectrum available, and the increasing flexibility of the FCC (Federal Communications Commission) in the field of spectrum regulation, has important policy implications for the rest of the world. The future existence of more unlicensed frequency could speed up developments leading towards a more fragmented industry structure with a rapid entry of new service providers.

In Asia, several countries are showing a desire to take the lead in 4G through ambitious, long-range plans and by aiming to achieve the early introduction of public standards for 4G systems. Korea and Japan are taking a proactive approach to the introduction of 4G. China is pursuing a leading role in 4G. In order to achieve this, the country has started developing its own technological standards such as TD-SCDMA (Time Division – Synchronous Code Division Multiple Access). It has also launched a number of government-sponsored research projects on 4G. Furthermore, a crucial step for China is the establishment of many joint ventures between Chinese and foreign companies, allowing Chinese companies to get both knowledge and capital. China’s large population, willingness to adopt new technologies and rapid economic growth means that 4G development there should be followed closely. If China succeeds in developing 4G systems, it can be anticipated that these will be offered at very competitive prices.

The main players in Asia are taking an entirely different approach by promoting a vision of a high data-rate public standard for the 4G system as a whole, building on strong demand for advanced data and entertainment services. Their 4G visions have many points in common with those of Europe, but on the whole, they tend to be more in line with the original linear vision of 4G developing as the next stage in the sequential evolution of mobile communications. They focus more on increasing mobile system data rates, and on developing new systems or system components, and less on the seamless operation of existing systems (though this latter strategy is increasingly included as the visions are further developed). These countries also envisage their governments taking an active role in driving domestic manufacturers to set early 4G standards.

1.1.5 TECHNOLOGY DEVELOPMENTS

The radio spectrum is a precious and scarce resource. Therefore, novel technologies for efficient spectrum utilisation to enhance the capacity of 3G and beyond systems are keenly anticipated. Factors that could have a significant impact on the deployment of mobile telecommunications technologies in this timeframe include radio access techniques enabling greater intelligence and flexibility to be built into transmitters and receivers. Some technology topics that appear relevant to some lesser or greater degree to the future development are: advanced radio resource management (RRM) algorithms; flexible frequency sharing methods; smart antennas; diversity techniques; coding techniques; space-time coding; efficient multiple access schemes or adaptive modulation.
Software Defined Radio (SDR) provides reconfigurable mobile communications systems that aim at providing a common platform to run software that addresses reconfigurable radio protocol stacks thereby increasing network and terminal capabilities and versatility through software modifications (downloads). Basically, SDR concerns all communication layers (from the physical layer to the application layer) of the radio interface and has an impact on both the user terminal and network side.

Future mobile user equipment may assume characteristics of general-purpose programmable platforms by containing high-power general-purpose processors and provide a flexible, programmable platform that can be applied to an ever-increasing variety of uses. The convergence of wireless connectivity and a general-purpose programmable platform might heighten some existing concerns and raise new ones; thus, environmental factors as well as traditional technology and market drivers influence the architecture of these devices. A well-designed embedded processor with a reconfigurable unit may enable user-defined instructions to be efficiently executed, since general-purpose processors such as CPUs or DSPs are not suitable for bit-level operation. This type of processor, which can handle many kinds of bit data processes, can be applied to various applications for mobile communication systems with efficient operation.

1.2 UMTS

3G mobile communications systems arose as a response to the challenge of developing systems that increased the capacity of the existing 2G systems. Simultaneously, they would provide a platform that allowed a seamless and ubiquitous access to the user of a wide range of new services, both circuit and packet switched, with higher requirements in terms of bit rate than those for which 2G systems were conceived. The development of 3G systems started in 1995, coordinated by the ITU-T (International Telecommunications Union – Telecommunications sector) under the generic terminology of IMT-2000 and so far different radio access technologies have been considered [6], leading to the development of several standards. Within this framework, the Universal Mobile Telecommunications System (UMTS) is the European proposal given by ETSI (European Telecommunications Standards Institute) to the 3G challenge. As a matter of fact, it is the dominant standard, resulting from the standardisation work done by the 3GPP (3rd Generation Partnership Project), an organisation formed by different regional standardisation bodies that include the presence of both manufacturers and operators from all around the world.

UMTS has been developed as the migration of the ETSI 2G/2.5G systems GSM/GPRS. The aim is to facilitate as much as possible the extension of the existing networks of these worldwide systems as well as the interoperability of the new UMTS system with the previous networks, thus allowing a progressive migration of the technology. As a result of this requirement, the most important changes introduced in the initial release of UMTS consist of a new radio access network based on a different radio access technology, while keeping the core network similar to that existing in GSM/GPRS systems. After this initial implementation, the subsequent releases of the UMTS system introduce important changes in the architecture of the core network, taking the Internet Protocol (IP) as the driving technology.

In the above context, this sub-section presents the main features of the UMTS network architecture, by defining the different elements that comprise it and that establish the basis over which Radio Resource Management (RRM) strategies, which are the main focus of this book, can be implemented.

1.2.1 UMTS ARCHITECTURE

The general UMTS network architecture from the physical point of view is presented in Figure 1.1 and it consists of an abstract model, applicable to any UMTS network, with independency of the specific release [7]. It is organised in domains, and each domain represents the highest level group of physical entities. Reference points are defined between the different domains. The basic split considers the User Equipment (UE) domain, used by the user to access the UMTS services, and the Infrastructure domain,
composed of the physical nodes, belonging to the network operator, that support the service requirements and the interconnection with the entity at the other end (e.g. another UE from the same or another network) with whom the end-to-end service has to be established. Both domains are separated by means of the Uu reference point, which represents the radio interface, and their elements are explained in the following sub-sections.

1.2.1.1 User Equipment Domain

The User Equipment domain consists of the terminal that allows the user access to the mobile services through the radio interface. From an architectural point of view, it is split into two sub-domains, separated by the Cu reference point (see Figure 1.1):

- **Mobile Equipment (ME) domain.** This represents the physical entity (e.g. a handset) that in turn is sub-divided into the Mobile Termination (MT) entity, which performs the radio transmission and reception, and the Terminal Equipment (TE), which contains the applications. These two entities may be physically located at the same hardware device depending on the specific application. For example, in the case of a handset used for a speech application, both MT and TE are usually located in the handset, while if the same handset is being used for a web browsing application, the handset will contain the MT and the TE can reside in, for example, an external laptop that contains the web browser.

- **UMTS Subscriber Identity Module (USIM) domain.** Typically, the physical hardware device containing the USIM is a removable smart card. The USIM contains the identification of the profile of a given user, including his identity in the network as well as information about the services that this user is allowed to access depending on the contractual relationship with the mobile network operator. So, the USIM is specific for each user and allows him/her to access the contracted services in a secure way by means of authentication and encryption procedures regardless of the ME that is used.

1.2.1.2 Infrastructure Domain

The infrastructure domain in the UMTS architecture contains the physical nodes that terminate the radio interface allowing the provision of the end-to-end service to the UE. In order to separate the
functionalities that are dependent on the radio access technology being used from those that are independent, the infrastructure domain is in turn split into two domains, namely the Access Network and the Core Network domains (see Figure 1.1), separated by the Iu reference point. This allows there to be a generic UMTS architecture that enables the combination of different approaches for the radio access technology as well as different approaches for the core network. As a matter of fact, notice that this architecture is the same used for GSM/GPRS networks so that the difference between a GSM/GPRS network and a UMTS network will mainly rely on the specific implementations of the access network and the core networks domains.

With respect to the core network, and in order to take into account different scenarios in which the user communicates with users in other types of networks (e.g. other mobile networks, fixed networks, Internet, etc.), three different sub-domains are defined (see Figure 1.1):

- **Home Network (HN) domain.** This corresponds to the network to which the user is subscribed, so it belongs to the operator that has the contractual relationship with the user. The user service profile as well as the user secure identification parameters are kept in the HN and should be coordinated with those included in the USIM at the UE.
- **Serving Network (SN) domain.** This represents the network containing the access network to which the user is connected in a given moment and it is responsible for transporting the user data from the source to the destination. Physically, it can be either the same HN or a different network in the case where the user is roaming with another network operator. The SN is then connected to the access network through the Iu reference point and to the HN through the [Zu] reference point. The interconnection with the HN is necessary in order to retrieve specific information about the user service abilities and for billing purposes.
- **Transit Network (TN) domain.** This is the core network part located on the communication path, between the SN and the remote party, and it is connected to the SN through the [Yu] reference point. Note that, where the remote party belongs to the same network to which the user is connected, the SN and the TN are physically the same network. Note also that, in general, the TN may not be a UMTS network, for example, in the case of a connection with a fixed network or when accessing the Internet.

According to this generic framework, several scenarios can be defined depending on the networks to which the UE and the remote party are connected, and it is even possible that the HN, the SN and the AN are physically the same network.

We will now describe the specific architectures of the AN for terrestrial UMTS networks, denoted as UTRAN (Universal Terrestrial Radio Access Network), and the UMTS generic Core Network (CN), which can be the HN, SN or TN.

**Universal Terrestrial Radio Access Network (UTRAN)** The architecture of the UTRAN is shown in Figure 1.2 [8]. It is composed of Radio Network Subsystems (RNSs) that are connected to the Core Network through the Iu interface that coincides with the Iu reference point of the overall UMTS architecture. Each RNS is responsible for the transmission and reception over a set of UMTS cells. The connection between the RNS and the UE is done through the Uu or radio interface.

The RNSs contain a number of Nodes B or base stations and one Radio Network Controller (RNC), connected through Iub interfaces. RNCs belonging to different RNSs are interconnected by means of the Iur interface.

A node B is the termination point between the air interface and the network and it is composed of one or several cells or sectors. In the 3GPP terminology, a cell stands as the smallest radio network entity that has its own identification number, denoted as Cell ID. Conceptually, a cell is regarded as a UTRAN Access Point through which radio links with the UEs are established. From a functional point of view, the cell executes the physical transmission and reception procedures over the radio interface.

The RNC is the node responsible for controlling the use of the radio resources in the nodes B that are under its control, thus it is the main entity where UMTS Radio Resource Management (RRM) algorithms
are executed. The majority of functionalities related to the radio interface are executed in the RNC, with the exception of the physical transmission and reception processes and some specific Medium Access Control (MAC) functions that are executed in the Node B. On the network side, the RNC interoperates with the CN through the Iu interface and establishes, maintains and releases the connections with the CN elements that the UEs under its control require in order to receive the UMTS services.

Additionally, as is shown in Figure 1.2, it is also possible for a RNC to interoperate with the Base Station Subsystems (BSSs) that form the GERAN (GSM/EDGE Radio Access Network) by means of the Iur-g interface. This interoperation allows the execution of Common Radio Resource Management (CRRM) algorithms between UMTS and GSM/GPRS systems.

From a functional point of view, the RNC may take several logical roles:

- **CRNC (Controlling RNC).** This is the role with respect to the Node B, and refers to the control that the RNC has over a set of Nodes B.
- **SRNC (Serving RNC).** This role is taken with respect to the UE. The SRNC is the RNC that holds the connection of a given UE with the CN through the Iu interface, so it can be regarded as the RNC that controls the RNS to which the mobile is connected at a given moment. When the UE moves across the network and executes handover between the different cells, it may require a SRNS (Serving RNS, i.e. the RNS having the SRNC) relocation procedure when the new cell belongs to a different RNC. This procedure requires the communication between the SRNC and the new RNC through the Iur interface in order for the new RNC to establish a new connection with the CN over its Iu interface.
- **DRNC (Drift RNC).** This role is also taken with respect to the UE and is a consequence of a specific type of handover that exists with CDMA systems, denoted as soft handover. In this case, a UE can be simultaneously connected to several cells (i.e. it has radio links with several cells). Then, when a UE moves in the border between RNSs, it is possible that it establishes new radio links with cells

![Figure 1.2 UTRAN architecture](image-url)
belonging to a new RNC while at the same time keeping the radio link with some cells of the SRNC. In this case, the new RNC takes the role of DRNC, and the connectivity with the CN is not done through the Iu of the DRNC but still through the Iu of the SRNC, thus requiring it to establish resources for the UE in the Iur interface between SRNC and DRNC. Only when all the radio links of the old RNC are released and the UE is connected only to the new RNC, will the SRNS relocation procedure be executed. Figure 1.3 illustrates the difference between SRNC and DRNC roles.

Notice that all the RNCs are CRNC and that a given RNC may be SRNC for certain UEs and simultaneously DRNC for others.

Two different operation modes have been standardised for the UTRAN radio interface and can be supported with the architecture of Figure 1.2 simply by changing the radio access technology. These modes are:

- **UTRAN FDD (Frequency Division Duplex) mode.** In this case, the uplink and downlink transmit with different carrier frequencies, thus requiring the allocation of paired bands. The access technique being used is WCDMA (Wideband Code Division Multiple Access), which means that several transmissions in the same frequency and time are supported and can be distinguished by using different code sequences.
- **UTRAN TDD (Time Division Duplex) mode.** In this case, the uplink and downlink operate with the same carrier frequency but in different time instants, thus they are able to use unpaired bands. The access technique being used is a combination of TDMA and CDMA, denoted as TD/CDMA, which means that simultaneous transmissions are distinguished by different code sequences (CDMA component) and that a frame structure is defined to allocate different transmission instants (time slots) to the different users (TDMA component).

The initial frequency bands reserved for each of the two UTRAN modes are shown in Figure 1.4 for regions 1 and 3 (i.e. Europa and Asia). The two radio access technologies lead to the existence of cells

<table>
<thead>
<tr>
<th>TDD</th>
<th>FDD (UL)</th>
<th>TDD</th>
<th>FDD (DL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1920 MHz</td>
<td>1980 MHz</td>
<td>2010 MHz</td>
<td>2025 MHz</td>
</tr>
<tr>
<td>2110 MHz</td>
<td>2170 MHz</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
supporting one or both of the two modes, as well as the ability to interoperate between them. Notice that, from the radio resource management point of view, the concept of radio resource is different for each mode. As a result, the RRM strategies in both cases lead to different types of algorithms. In the context of this book, only the RRM strategies for the UTRAN FDD mode are considered.

**Core Network** The Core Network (CN) is the part of the mobile network infrastructure that covers all the functionalities that are not directly related with the radio access technology, thus it is possible to combine different core network architectures with different radio access networks. Examples of these functionalities are the connection and session management (i.e. establishment, maintenance and release of the connections and sessions for circuit switched and packet switched services) as well as mobility management (i.e. keep track of the area where each UE can be found in order to route calls to it).

While the access network in UMTS has suffered relatively few changes since the initial UMTS release (release 99), this is not the case with the Core Network. The reason is that the initial implementations of UMTS were seen simply as an extension of the GSM/GPRS networks because they maintained the existing core network for GSM/GPRS (with small modifications) in order to make it compatible with the new UMTS access network. This was due to the impression that the new radio interface technology posed the most critical challenges in the support of the expected UMTS services. The new releases of UMTS introduced the major changes in the architecture of the core network only after the radio access part was stabilised, therefore driving it towards the development of an all IP network.

Figure 1.5 shows the elements that compose the architecture of the UMTS core network as well as the interfaces between them [9]. The figure reflects the UMTS release 99, which is essentially the same as the GSM/GPRS system, and the evolution of this architecture in future releases will be explained in Section 1.2.2. As can be observed, the infrastructure of the core network is divided into two domains that differ in the way they support user traffic. They are the Circuit Switched (CS) and the Packet Switched (PS) domains. The CS domain supports the traffic composed by connections that require dedicated

![Figure 1.5 Architecture of the UMTS Core Network (Release 99)](image-url)
network resources, and allows the interconnection with external CS networks like the PSTN (Public Switched Telephone Network) or the ISDN (Integrated Services Digital Network). In turn, the PS domain supports a traffic composed of packets, which are groups of bits that are autonomously transmitted and independently routed, so no dedicated resources are required throughout the connection time, since the resources are allocated on a packet basis only when needed. This allows a group of packet flows to share the network resources based on traffic multiplexing. The PS domain allows the interconnection of external PS networks, like the Internet. The division between CS and PS domains introduces the requirement to split the Iu reference point between core and access networks in two interfaces, denoted as Iu_CS and Iu_PS.

There are some entities in the CN that belong both to the CS and the PS domains. They are the HLR (Home Location Register), the AuC (Authentication Centre) and the EIR (Equipment Identity Register). The HLR is a database that stores information about the users that are subscribed in a given network, including the different user identifiers and the service profile. The AuC stores the identity keys of the subscribed users and is used by the HLR to perform security operations. In turn, the EIR is a database that stores the identifiers of the mobile terminals in order to detect those terminals whose access to the network must be denied due to different reasons (for example, because they have been stolen).

The CS domain is composed of three specific entities, namely the MSC (Mobile Switching Centre), the GMSC (Gateway Mobile Switching Centre) and the VLR (Visitor Location Register). The MSC interacts with the radio access network by means of the Iu_CS interface and executes the necessary operations to handle CS services. This includes routing the calls towards the corresponding transit network and establishing the corresponding circuits in the path. The MSC is the same as that which uses the GSM network with the difference that a specific interworking function is required between the MSC and the access network in UMTS. The reason is that in GSM the speech traffic delivered to the core network by the access network uses 64 kb/s circuits while in UMTS the speech uses adaptive multi-rate technique (AMR) with bit rates between 4.75 kb/s and 12.2 kb/s that are transported in the access network with ATM (Asynchronous Transfer Mode) technology [6]. This is why the term 3G MSC is sometimes used to differentiate between the MSC from GSM system and the MSC from UMTS networks.

The VLR is a database associated with a MSC that contains specific information (e.g. identifiers, location information, etc.) about the users that are currently in the area of this MSC, which allows the performing of certain operations without the need to interact with the HLR. The information contained in the VLR and the HLR must be coordinated.

The GMSC is a specific MSC that interfaces with the external CS networks and is responsible of routing calls to/from the external network. To this end, it interacts with the HLR to determine the MSC through which the call should be routed. In release 99, the communication between the entities of the CS domain is done by means of 64 kb/s circuits and uses SS7 (Signalling System No. 7) for signalling purposes.

The PS domain is composed of two specific entities, namely the SGSN (Serving GPRS Support Node) and GGSN (Gateway GPRS Support Node), which perform the necessary functions to handle packet transmission to and from the UEs. The SGSN is the node that serves the UE and establishes a mobility management context including security and mobility information. It interacts with the UTRAN by means of the Iu_PS interface. The GGSN, in turn, interfaces with the external data networks and contains routing information of the attached users. IP tunnels between the GGSN and the SGSN are used to transmit the data packets of the different users [10].

### 1.2.2 UMTS EVOLUTION

The development of the first UMTS specifications was done at a time when the Internet was becoming progressively more and more popular and the IP technology began to be used not only for the transport of data services but also for speech and video services, thus becoming a new paradigm for the deployment of multiservice networks. The response of the UMTS system to this expansion of IP technology is given
in the releases that followed the initial release 99, and whose main purpose was the progressive transformation of UMTS in an all IP network that was more efficient than the coexistence of two separated networks for the CS and PS core network domains.

Figure 1.6 tries to show in a schematic way the main changes, represented with black stars, which appear in the different UMTS releases. In release 99, the transport technology used between the elements of the UTRAN is ATM. In turn, in the CS domain of the core network, 64 kb/s circuits are used and in the PS domain transmissions are done by means of IP tunnels using GTP (GPRS Tunnelling Protocol). The first step in the evolution towards an all IP network is release 4, in which the CS domain of the CN is replaced by an IP or an IP/ATM backbone. The transmission of speech services over this IP backbone introduces important technological challenges that lead to the so-called voice over IP technology. This modification of the core network involves the evolution of MSC in two different components, namely the MSC server, which comprises the call control and mobility control parts of the MSC, thus handling only signalling, and the MGW (Media Gateway function), which handles the users' data flows.

Release 5 executes the final step to achieving a CN completely based on IP technology by removing the possibility of using ATM in the CS domain. In the new architecture, the provision of real time IP multimedia service is done by means of the inclusion of a new CN domain, namely the IP Multimedia Subsystem (IMS), which is connected to the GGSN and the MGW, and makes use of the Session Initiation Protocol (SIP) as a means of establishing multimedia sessions between users supporting user mobility and call redirection [11]. Another of the changes introduced by release 5 in the CN consists of the integration of the functionalities of the HLR and the AuC in the HSS (Home Subscriber Server), which contains the subscription related information for each user in order to support the call and session handling.

The release 5 does not limit the changes to the CN, and introduces an important modification at the radio interface as well. In particular, a new packet access mechanism over WCDMA, denoted as HSDPA (High Speed Downlink Packet Access) is defined. HSDPA supports much higher bit rates up to around 10 Mb/s by means of an additional modulation scheme and the implementation of fast packet scheduling and hybrid retransmission mechanisms, and coexists with the radio access mechanisms existing in previous releases [12].
The final objective of an all IP architecture like the one defined in release 5 is the inclusion of the IP technology in the radio access network as well. Due to the important modifications that such a change requires, this inclusion was postponed to the release 6. The existence of an all IP network including the radio access part facilitates the integration of different radio access technologies operating over a unique backbone technology and therefore enables the development of heterogeneous networks that integrate the UTRAN and GERAN technologies with others like WLAN.

1.3 QoS MODEL IN UMTS

In order to provide a service with specific QoS requirements, a bearer service with clearly defined characteristics and functionality needs to be set up from the source to the destination of the service. Since the end-to-end path extends across different system levels each having their own QoS properties, the QoS is handled and split in different parts taking into account the special characteristics of each component. In this framework, the UMTS QoS mechanisms shall provide a mapping between application requirements and UMTS services.

The layered architecture of a UMTS bearer service is depicted in Figure 1.7. Each bearer service on a specific layer offers its individual services using the services provided by the layers below. The end-to-end service used by the TE (Terminal Equipment) will be realised using a TE/MT Local Bearer Service, a UMTS Bearer Service, and an External Bearer Service. The QoS mechanisms outside the UMTS network are not within the scope of 3GPP specifications and, consequently, the end-to-end bearer service is beyond the scope of specification TS 23.107 [13], which is mainly described in this section. Nevertheless, it is worth noting that the UMTS operator offers a wide variety of services by means of the UMTS Bearer Service. In turn, the UMTS Bearer Service consists of two parts, the Radio Access Bearer (RAB) Service and the Core Network Bearer Service. In this way, an optimised realisation of the UMTS Bearer Service over the respective segments is more feasible. The Radio Access Bearer Service is based

![QoS model in UMTS](image-url)
on the characteristics of the radio interface and is maintained for a moving MT. The role of the Core Network Bearer Service is to control and utilise efficiently the backbone network.

The Radio Access Bearer Service is realised by a Radio Bearer Service and an Iu-Bearer Service. The role of the Radio Bearer Service is to cover all the aspects of the radio interface transport. In the context of this book, it is considered that this bearer service uses the UTRAN FDD. The Iu-Bearer Service together with the Physical Bearer Service provides the transport between UTRAN and CN.

In UMTS, four different QoS classes have been identified:

- **Conversational class.** The Real time conversation scheme is characterised by a low transfer time because of the conversational nature of the scheme and fact that the time variation between information entities of the stream will be preserved in the same way as for real time streams. The maximum transfer delay is given by the human perception of video and audio conversation. The most well known use of this scheme is telephony speech. Nevertheless, with the Internet and multimedia, a number of new applications will require this scheme, for example voice over IP (VoIP) and video conferencing tools.

- **Streaming class.** This scheme is one of the newcomers in data communication, raising a number of new requirements in telecommunication systems. It is characterised by the fact that the time variation between information entities (i.e. samples, packets) within a flow will be preserved, although it does not have any requirements on low transfer delay. As the stream normally is time aligned at the receiving end (in the user equipment), the highest acceptable delay variation over the transmission media is given by the capability of the time alignment function of the application. Acceptable delay variation is thus much greater than the delay variation given by the limits of human perception.

- **Interactive class.** Interactive traffic is the other classical data communication scheme that on an overall level is characterised by the request response pattern of the end user. This scheme applies when the end user, which can be either a machine or a human, is online requesting data from remote equipment (e.g. a server). Examples of human interaction with the remote equipment are: web browsing, data base retrieval, server access. Examples of machines interaction with remote equipment are: polling for measurement records and automatic database enquiries (tele-machines). At the message destination, there is an entity expecting the response within a certain time. Round trip delay time is therefore one of the key attributes. Another characteristic is that the content of the packets are transparently transferred (i.e. with low bit error rate).

- **Background class.** When the end user, which typically is a computer, sends and receives data-files in the background, this scheme applies. Examples are background delivery of emails, SMS (Short Message Service), download of databases and reception of measurement records. Background traffic is one of the classical data communication schemes that on an overall level is characterised by the fact that the destination is not expecting the data within a certain time. The scheme is thus more or less delivery time insensitive. Another characteristic is that the content of the packets are transparently transferred (i.e. with low bit error rate).

The Radio Access Bearer Service attributes, which will be applied to both CS and PS domains, are:

- **Traffic class** (‘conversational’, ‘streaming’, ‘interactive’, ‘background’). With this attribute, UTRAN can make assumptions about the traffic source and optimise the transport for that traffic type.

- **Maximum bit rate.** This is the maximum number of bits delivered by UTRAN or to UTRAN at a SAP (Service Access Point) within a period of time, divided by the duration of the period. The purpose of this attribute is mainly to limit the delivered bit rate to applications or external networks as well as to allow the maximum desired RAB bit rate to be defined for applications able to operate with different bit rates.

- **Guaranteed bit rate.** This is the guaranteed number of bits delivered at a SAP within a period of time (provided that there are data to deliver), divided by the duration of the period. This attribute may be used to facilitate admission control based on available resources, and for resource allocation within
UTRAN. Quality requirements expressed by, for example, delay and reliability attributes, only apply to incoming traffic up to the guaranteed bit rate. It is worth noting that the guaranteed bit rate at the RAB level may be different from that on the UMTS bearer level, for example due to header compression.

- Delivery order. This indicates whether the UMTS bearer shall provide in-sequence SDU (Service Data Unit) delivery or not and specifies if out-of-sequence SDUs are acceptable or not.
- Maximum SDU size used for admission control and policing. This corresponds to the maximum packet size that can be delivered at the top of the radio interface.
- SDU format information. This is the list of possible exact sizes of SDUs.
- SDU error ratio. This indicates the fraction of SDUs lost or detected as erroneous. This attribute is used to configure the protocols, algorithms and error detection schemes, primarily within UTRAN.
- Residual bit error ratio. This indicates the undetected bit error ratio in the delivered SDUs. It is used to configure radio interface protocols, algorithms and error detection coding.
- Delivery of erroneous SDUs. This indicates whether SDUs detected as erroneous will be delivered or discarded.
- Transfer delay. This indicates the maximum delay for the 95th percentile of the distribution of delay for all delivered SDUs during the lifetime of a bearer service, where delay of an SDU is defined as the time from a request to transfer an SDU at one SAP to its delivery at the other SAP. The attribute is used to specify the delay tolerated by the application and allows UTRAN to set transport formats and ARQ (Automatic Repeat Request) parameters.
- Traffic handling priority, specifying the relative importance of handling all SDUs belonging to the UMTS bearer compared to the SDUs of other bearers. In particular, there is a need to differentiate between bearer qualities within the interactive class. This is handled with this attribute, to allow UMTS to schedule traffic accordingly. By definition, priority is an alternative to absolute guarantees, and thus these two attributes cannot be used together for a single bearer.
- Allocation/Retention Priority. This specifies the relative importance, compared to other UMTS bearers, of allocation and retention of the UMTS bearer. In situations where resources are scarce, this attribute may be used to prioritise bearers when performing admission control.
- Source statistics descriptor. This specifies characteristics of the source of submitted SDUs and it may take the values ‘speech’ or ‘unknown’. Since conversational speech has a well-known statistical behaviour, UTRAN may calculate a statistical multiplex gain for use in admission control on the radio and Iu interfaces.

It is worth remarking that, when establishing a UMTS bearer and the underlying Radio Access Bearer for support of a service request, some attributes typically have different values on both levels. For example, the requested transfer delay of the UMTS bearer will typically be larger than the requested transfer delay of the Radio Access Bearer, as the transport through the core network will use part of the acceptable delay. Similarly, SDU error ratio for Radio Access Bearer service will be reduced with the errors introduced in the core network, by the Core Network Bearer service. Furthermore, some attributes/ settings only exist on the Radio Access Bearer level, such as Source statistics descriptor.

REFERENCES

Introduction