1

Introduction

Ari Ahtianen, Heikki Kaaranen and Siamäk Naghian

Nowadays, it is widely recognised that there are three different, implemented generations as far as mobile communication is concerned (Figure 1.1). The first generation, 1G, is the name for the analogue or semi-analogue (analogue radio path, but digital switching) mobile networks established in the mid-1980s, such as the Nordic Mobile Telephone (NMT) system and the American Mobile Phone System (AMPS). These networks offered basic services for users and the emphasis was on speech and speech-related services. 1G networks were developed with national scope only and very often the main technical requirements were agreed between the governmental telecom operator and the domestic industry without wider publication of the specifications. Due to national specifications, 1G networks were incompatible with each other and mobile communication was considered at that time to be some kind of curiosity and added value service on top of the fixed networks.

Because the need for mobile communication increased, also the need for a more global mobile communication system arose. International specification bodies started to specify what the second generation, 2G, mobile communication system should look like. The emphasis for 2G was on compatibility and international transparency; the system should be regional (e.g., European-wide) or semi-global and the users of the system should be able to access it basically anywhere within the region. From the end-user's point of view, 2G networks offered a more attractive "package" to buy; besides the traditional speech service these networks were able to provide some data services and more sophisticated supplementary services. Due to the regional nature of standardisation, the concept of globalisation did not succeed completely and there are some 2G systems available on the market. Of these, the commercial success story is the Global System for Mobile Communications (GSM) and its adaptations: it has clearly exceeded all the expectations set, both technically and commercially.

The third generation, 3G, is expected to complete the globalisation process of mobile communication. Again, there are national and regional interests involved and difficulties can be foreseen. Anyway, the trend is that 3G will mostly be based on GSM technical solutions for two reasons: GSM technology dominates the market and the great investments made in GSM should be utilised as much as possible. Based on this, the specification bodies created a vision about how mobile telecommunication will

UMTS Networks Second Edition H. Kaaranen, A. Ahtiainen, L. Laitinen, S. Naghian and V. Niemi
 ISBN: 0-470-01103-3

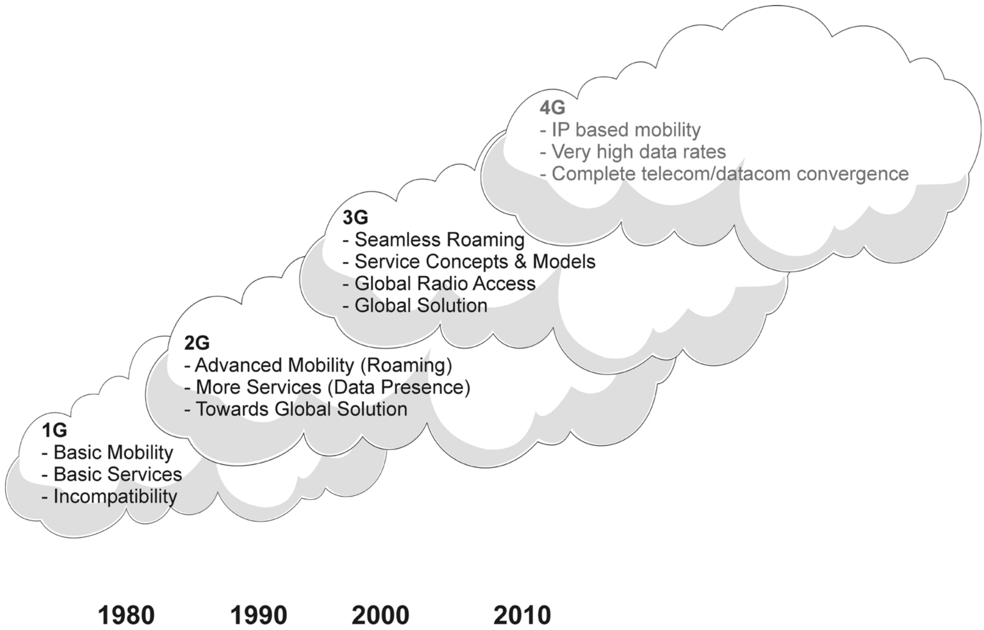

Figure 1.1 Cellular generations

develop within the next decade. Through this vision, some requirements for 3G were shortlisted as follows:

1. The system must be fully specified (like GSM) and major interfaces should be standardised and open. The specifications generated should be valid worldwide.
2. The system must bring clear added value to GSM in all aspects. However, at the start the system must be backward-compatible at least with GSM and ISDN (Integrated Services Digital Network).
3. Multimedia and all of its components must be supported throughout the system.
4. The radio access of 3G must provide wideband capacity that is generic enough to become available worldwide. The term "wideband" was adopted to reflect the capacity requirements between 2G narrowband capacity and the broadband capacity of fixed communications media.
5. The services for end-users must be independent of radio access technology details and the network infrastructure must not limit the services to be generated. That is, the technology platform is one issue and the services using the platform are totally another issue.

While 3G specification work was still going on, the major telecommunication trends changed too. The traditional telecommunication world and up to now the separate data communications (or the Internet) have started to converge rapidly. This has started a development chain, where traditional telecommunication and Internet Protocol (IP) technologies are combined in the same package. This common trend has many names depending on the speaker's point of view; some people call the target of this development the "Mobile Information Society" or "Mobile IP", others say it is "3G All IP" and in some commercial contexts the name "E2E IP" (End-to-End IP) is used as well. From a 3G point of view, a full-scale IP implementation is defined as a single targeted phase of the 3G development path.

The 3G system experiences evolution through new phases and, actually, the work aiming to establish 4G specifications has already started. Right now it may be too early to predict where the 3G evolution ends and 4G really starts. Rather, this future development can be thought of as an ongoing development chain where 3G will continue to introduce new ways of handling and combining all kinds of data and mobility. 4G will then emerge as a more sophisticated system concept bringing still more capacity and added value to end-users.

1.1 Specification Process for 3G

The uniform GSM standard in European countries has enabled globalisation of mobile communications. This became evident when the Japanese 2G Pacific Digital Communications (PDC) failed to spread to the Far East and the open GSM standard was adopted by major parts of the Asian markets and when its variant became one of the nationally standardised alternatives for the US Personal Communication System (PCS) market too.

A common, global mobile communication system naturally creates a lot of political desires. In the case of 3G this can be seen even in the naming policy of the system. The most neutral term is "third generation", 3G. In different parts of the world different issues are emphasised and, thus, the global term 3G has regional synonyms. In Europe 3G has become UMTS (Universal Mobile Telecommunication System), following the European Telecommunications Institute (ETSI) perspective. In Japan and the US the 3G system often carries the name IMT-2000 (International Mobile Telephony 2000). This name comes from the International Telecommunication Union (ITU) development project. In the US the CDMA2000 (Code Division Multiple Access) is also an aspect of 3G cellular systems and represents the evolution from the IS-95 system. In this book, we will describe the UMTS system as it has been specified by the worldwide 3G Partnership Project (3GPP). To bring some order to the somewhat confusing naming policy, 3GPP launched a decision where it stated that the official name of 3G is the "3GPP System". This name should be followed by a release number describing the specification collection. With this logic, the very first version of the European-style UMTS network takes the official name "3GPP System Release 99". Despite this definition, the above-mentioned names UMTS and IMT-2000 are still widely used.

At the outset UMTS inherited plenty of elements and functional principles from GSM and the most considerable new development is related to the radio access part of the network. UMTS brings into the system an advanced access technology (namely, the wideband type of radio access). Wideband radio access is implemented using Wideband Code Division Multiple Access (WCDMA) technology. WCDMA evolved from CDMA, which, as a proven technology, has been used for military purposes and for narrowband cellular networks, especially in the US.

UMTS standardisation was preceded by several pre-standardisation research projects founded and financed by the EU. Between 1992 and 1995 a Research in Advanced Communications in Europe (RACE) MoNet project developed the modelling technique describing the function allocation between the radio access and core parts of the network. This kind of modelling technique was needed, for example, to compare Intelligent Network (IN) and GSM Mobile Application Part (MAP) protocols as mobility management solutions. This was, besides the discussion on the broadband versus narrowband ISDN, one of the main dissents in MoNet. In addition, discussions about the use of ATM (Asynchronous Transfer Mode) and B-ISDN as fixed transmission techniques arose at the end of the MoNet project.

Between 1995 and 1998 3G research activities continued within the Advanced Communications Technology and Services (ACTS) Future Radio Wideband Multiple Access System (FRAMES) project. The first years were used for selecting and developing a suitable multiple access technology, considering mainly the TDMA (Time Division Multiple Access) versus CDMA. The big European manufacturers preferred TDMA because it was used also in GSM. CDMA-based technology was promoted mainly by US industry, which had experience with this technology mainly due to its early utilisation in defence applications.

ITU dreamed of specifying at least one common global radio interface technology. This kind of harmonisation work was done under the name "Future Public Land Mobile Telephony System" (FPLMTS) and later IMT-2000. Due to many parallel activities in regional standardisation bodies this effort turned into a promotion of common architectural principles among the family of IMT-2000 systems.

Europe and Japan also had different short-term targets for 3G system development. In Europe a need for commercial mobile data services with guaranteed quality (e.g., mobile video services) was widely recognised after the early experiences from narrowband GSM data applications. Meanwhile, in the densely populated Far East there was an urgent demand for additional radio frequencies for speech services. The frequency bands identified by ITU in 1992 for the future 3G system called "IMT-2000" became the most obvious solution to this issue. In early 1998 a major push forward was achieved when ETSI TC-SMG decided to select WCDMA as its UMTS radio technology. This was also supported by the largest Japanese operator NTT DoCoMo. The core network technology was at the same time agreed to be developed on the basis of GSM core network technology. During 1998 the European ETSI and the Japanese standardisation bodies (TTC and ARIB) agreed to make a common UMTS standard. After this agreement, the 3GPP organisation was established and the determined UMTS standardisation was started worldwide.

From the UMTS point of view, the 3GPP organisation is a kind of "umbrella" aiming to form compromised standards by taking into account political, industrial and commercial pressures coming from the local specification bodies:

- ETSI (European Telecommunication Standard Institute)/Europe.
- ARIB (Association of Radio Industries and Business)/Japan.
- CWTS (China Wireless Telecommunication Standard group)/China.
- T1 (Standardisation Committee T1—Telecommunications)/US.

- TTA (Telecommunication Technology Association)/Korea.
- TTC (Telecommunications Technology Committee)/Japan.

As this is a very difficult task an independent organisation called the "OHG" (Operator Harmonisation Group) was established immediately after the 3GPP was formed. The main task for 3GPP is to define and maintain UMTS specifications, while the role of OHG is to look for compromise solutions for those items the 3GPP cannot handle internally. This arrangement guarantees that 3GPP's work will proceed on schedule.

To ensure that the American viewpoint will be taken into account a separate 3GPP Number 2 (3GPP2) was founded and this organisation performs specification work from the IS-95 radio technology basis. The common goal for 3GPP, OHG and 3GPP2 is to create specifications according to which a global cellular system having wideband radio access could be implemented. To summarise, there were three different approaches towards the global cellular system, 3G. These approaches and their building blocks are, on a rough level, presented in Table 1.1.

When globality becomes a reality, the 3G specification makes it possible to take any of the switching systems mentioned in the table and combine them with any of the specified radio access parts and the result is a functioning 3G cellular network. The second row represents the European approach known as "UMTS" and this book gives an overview of its first release.

The 3GPP originally decided to prepare specifications on a yearly basis, the first specification release being Release 99. This first specification set has a relatively strong "GSM presence". From the UMTS point of view the GSM presence is very important; first, the UMTS network must be backward-compatible with existing GSM networks and, second, GSM and UMTS networks must be able to interoperate together. The next release was originally known as "3GPP R00", but, because of the multiplicity of changes proposed, specification activities were scheduled into two specification releases 3GPP R4 and 3GPP R5. 3GPP R4 defines optional changes in the UMTS core network circuit-switched side; these are related to the separation of user data flows and their control mechanisms. 3GPP R5 aims to introduce a UMTS network providing mechanisms and arrangements for multimedia. This entity is known as the "IP Multimedia Subsystem" (IMS) and its architecture is presented in Chapter 6. IP and the overlying protocols will be used in network control too and user data

Table 1.1 3G variants and their building blocks

Variant	Radio access	Switching	2G basis
3G (US)	WCDMA, EDGE, CDMA2000	IS-41	IS-95, GSM1900, TDMA
3G (Europe)	WCDMA, GSM, EDGE	Advanced GSM NSS and packet core	GSM900/1800
3G (Japan)	WCDMA	Advanced GSM NSS and packet core	PDC

flows are expected to be mainly IP-based as well. In other words, the mobile network implemented according to the 3GPP R5 specification will be an end-to-end packet-switched cellular network using IP as the transport protocol instead of SS7 (Signalling System #7), which holds the major position in existing circuit-switched networks. Naturally, the IP-based network should still support circuit-switched services too. 3GPP R4/R5 will also start to utilise the possibility of new radio access techniques. In 3GPP R99 the basis for the UMTS Terrestrial Access Network (UTRAN) is WCDMA radio access. In 3GPP R4/5 another radio access technology derived from GSM with Enhanced Data for GSM Evolution (EDGE) is integrated to the system in order to create the GSM/EDGE Radio Access Network (GERAN) as an alternative to building a UMTS mobile network.

1.2 Introduction to the 3G Network Architecture

The main idea behind 3G is to prepare a universal infrastructure able to carry existing and also future services. The infrastructure should be designed so that technology changes and evolution can be adapted to the network without causing uncertainties in the existing services using the current network structure. Separation of access technology, transport technology, service technology (connection control) and user applications from each other can handle this very demanding requirement. The structure of a 3G network can be modelled in many ways, and here we introduce some ways to outline the basic structure of the network. The architectural approaches to be discussed in this section are:

- Conceptual network model.
- Structural network architecture.
- Resource management architecture.
- UMTS bearer architecture.

1.2.1 Conceptual Network Model

From the above-mentioned network conceptual model point of view, the entire network architecture can be divided into subsystems based on the nature of traffic, protocol structures and physical elements. As far as the nature of traffic is concerned, the 3G network consists of two main domains, packet-switched (PS) and circuit-switched (CS) domains. According to 3GPP specification TR 21.905 a *domain* refers to the highest level group of physical entities and the defined interfaces (reference points) between such domains. The interfaces and their definitions describe exactly how the domains communicate with each other.

From the protocol structure and their responsibility point of view, the 3G network can be divided into two strata: the access stratum and the non-access stratum. A *stratum* refers to the way of grouping protocols related to one aspect of the services provided by one or several domains (see 3GPP specification TR 21.905). Thus, the access stratum contains the protocols that handle activities between the User Equipment (UE) and the access network. The non-access stratum contains the protocols that

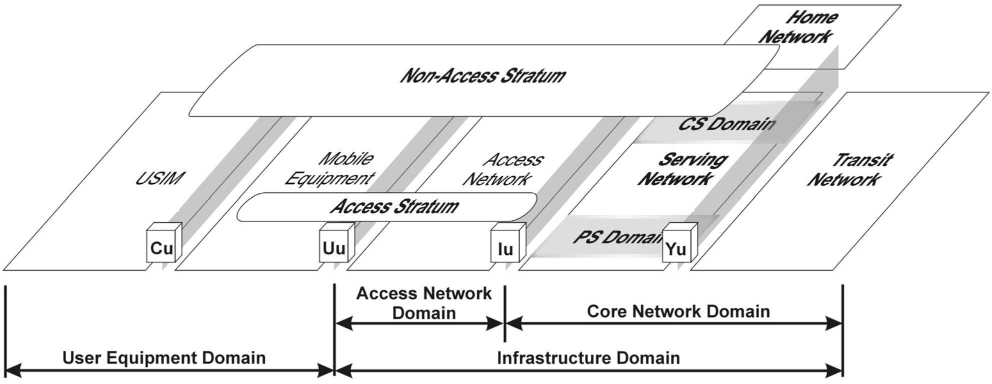

Figure 1.2 UMTS architecture—conceptual model

handle activities between the UE and the core network (CS/PS domain), respectively. For further information about strata and protocols see Chapter 10.

The part of Figure 1.2 called "Home Network" maintains static subscription and security information. The serving network is the part of the core network + domain which provides the core network functions locally to the user. The transit network is the core network part located on the communication path between the serving network and the remote party. If, for a given call, the remote party is located inside the same network as the originating UE, then no particular instance of the transit network is needed.

1.2.2 Structural Network Architecture

In this book we mainly present the issues from the network structural architecture perspective. This perspective is presented in Figure 1.3. In UMTS the GSM technology plays the remarkable role of the background and, actually, UMTS aims to reuse everything, which is reasonable. For example, some procedures used within the non-access stratum are, in principle, reused from GSM but naturally with required modifications.

The 3G system terminal is called "UE" and it contains two separate parts, Mobile Equipment (ME) and the UMTS Service Identity Module (USIM).

The new subsystem controlling wideband radio access has different names, depending on the type of radio technology used. The general term is "Radio Access Network" (RAN). When we talk in particular about UMTS with WCDMA radio access, the name "UTRAN" or "UTRA" is used. The other type of RAN included in UMTS is GERAN. GERAN and its definitions are not part of 3GPP R99, though they are referred to as possible radio access alternatives, which may be utilised in the future. The specification of GERAN and its harmonisation with UTRAN is done in 3GPP R4 and 3GPP R5.

UTRAN is divided into Radio Network Subsystems (RNSs). One RNS consists of a set of radio elements and their corresponding controlling element. In UTRAN the radio element is Node B, referred to as Base Station (BS) in the rest of this book, and the controlling element is the Radio Network Controller (RNC). The RNSs are connected

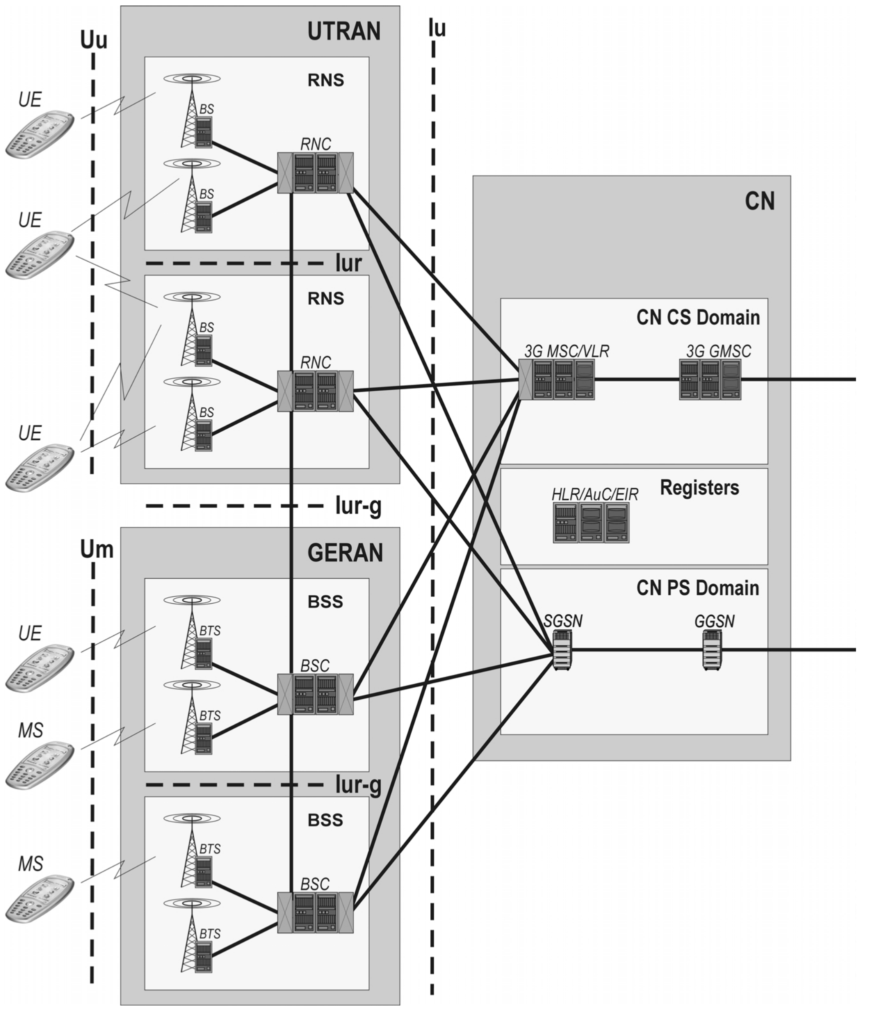

Figure 1.3 UMTS network architecture—network elements and their connections for user data transfer

to each other over the access network internal interface Iur. This structure and its advantages are explained in more detail in Chapter 5.

The other access network shown in Figure 1.3, GERAN, is not handled in detail in this book. Readers interested in GERAN should consult, e.g., Halonen et al. (2002).

The term ‘‘Core Network’’ (CN) covers all the network elements needed for switching and subscriber control. In early phases of UMTS, part of these elements were directly inherited from GSM and modified for UMTS purposes. Later on, when transport technology changes, the core network internal structure will also change in a remarkable way. CN covers the CS and PS domains defined in Figure 1.3. Configuration alternatives and elements of the UMTS core network are discussed in Chapter 6.

The part of Figure 1.3 called "Registers" is the same as the Home Network in the preceding 3G network conceptual model. This part of the network maintains static subscription and security information. Registers are discussed in more detail in Chapter 6.

The major open interfaces of UMTS are also presented in Figure 1.3. Between the UE and UTRAN the open interface is Uu, which in UMTS is physically realised with WCDMA technology. Some additional information about WCDMA on a general level is provided in Chapters 3 and 4. On the GERAN side the equivalent open interface is Um. The other major open interface is Iu located between UTRAN/GERAN and CN.

The RNSs are separated from each other by an open interface Iur. Iur is a remarkable difference when compared with GSM; it brings completely new abilities for the system to utilise: so-called macro diversity as well as efficient radio resource management and mobility mechanisms. When the Iur interface is implemented in the network, the UE may attach to the network through several RNCs, each of which maintains a certain logical role during radio connection. These roles are Serving RNC (SRNC), Drifting RNC (DRNC) and Controlling RNC (CRNC). CRNC has overall control of the logical resources of its UTRAN access points, being mainly BSs. An SRNC is a role an RNC can play with respect to a specific connection between the UE and UTRAN. There is one SRNC for each UE that has a radio connection to UTRAN. The SRNC is in charge of the radio connection between the UE and the UTRAN. It also maintains the Iu interface to the CN, which is the main characteristic of the SRNC. A DRNC plays the logical role used when radio resources of the connection between the UTRAN and the UE need to use cell(s) controlled by another RNC rather than the SRNC itself. UTRAN-related issues in general are discussed in Chapter 5.

Access networks also have connections between themselves through an interface Iur-g. Iur-g is used for radio-resource-management-related information transfer. The difference between Iur and Iur-g is that Iur transfers both signalling and user data, while Iur-g only transfers signalling.

In addition to the CS and PS domains presented in Figure 1.3, the network may contain other domains. One example of these is the broadcast messaging domain, which is responsible for multicast messaging control. However, in this book we concentrate on the UMTS network as presented in Figure 1.4. As far as the various RANs are concerned, we concentrate on UTRAN and highlight some specific items related on UTRAN–GERAN co-existence and co-operation.

1.2.3 Resource Management Architecture

The network element-centric architecture described above results from functional decomposition and the split of responsibilities between major domains and, ultimately, between network elements. Figure 1.4 illustrates this split of major functionalities, which are:

- Communication Management (CM).
- Mobility Management (MM).
- Radio Resource Management (RRM).

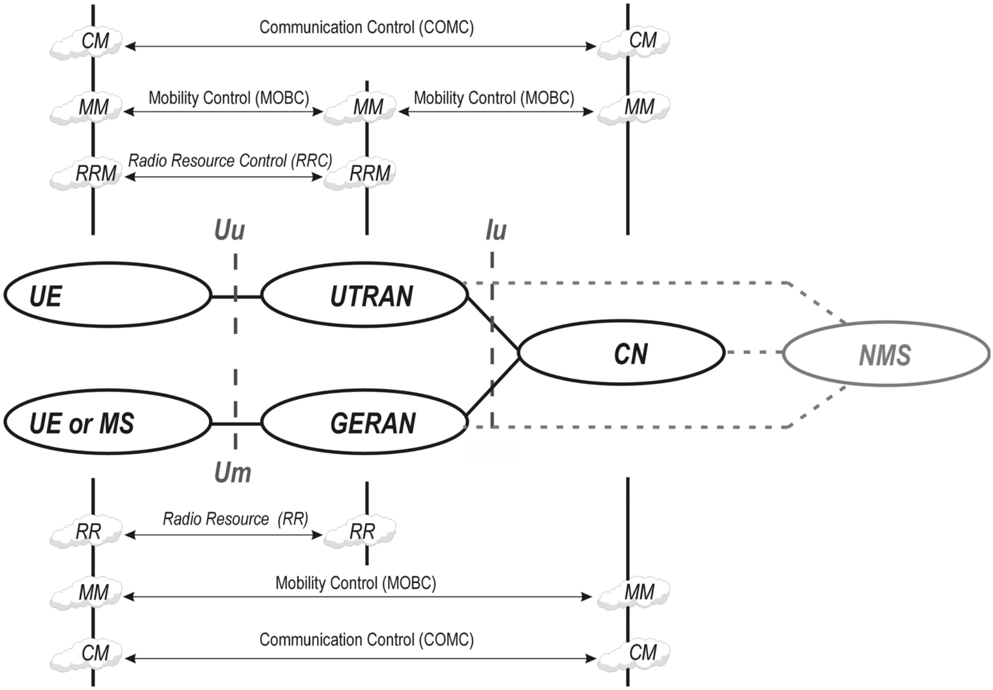

Figure 1.4 UMTS network architecture—management tasks and control duties

CM covers all of the functions and procedures related to the management of user connections. CM is divided into several sub-areas, such as call handling for CS connections, session management for PS connections, as well as handling of supplementary services and short-message services. MM covers all of the functions and procedures needed for mobility and security (e.g., connection security procedures and location update procedures). Most of the MM procedures occur within the CN and its elements, but in the 3G part of the MM functions are also performed in UTRAN for PS connections. The principles underlying CM and MM are discussed in Chapter 6.

RRM is a collection of algorithms UTRAN uses for management of radio resources. These algorithms handle, for instance, the power control for radio connections, different types of handovers, system load and admission control. RRM is an integral part of UTRAN and basic RRM is discussed more closely in Chapter 5. Some system-wide procedure examples about CM, MM and RRM functioning are given in Chapter 11.

Although these management tasks can be located within specific domains and network elements, they need to be supported by communication among the related domains and network elements. This communication is about gathering information and reporting about the status of remote entities as well as about giving commands to them in order to execute management decisions. Therefore, each of the management tasks is associated with a set of control duties such as:

- Communication Control (COMC).
- Mobility Control (MOBC).
- Radio Resource Control (RRC).

COMC maintains mechanisms like call control and packet session control. MOBC maintains mechanisms which cover, for example, execution control for location updates and security. Radio resources are completely handled between UTRAN and the UE. The control duty called "RRC" takes care of, for example, radio link establishment and maintenance between UTRAN and the UE. These collections of control duties are then further refined into a set of well-specified control protocols. For more detailed information about protocols see Chapter 10.

When compared with the traditional GSM system, it is apparent that this functional architecture has undergone some rethinking. The most visible change has to do with mobility management, where responsibility has been split between UTRAN and the CN. In addition, with regards to the RRM, the UMTS architecture follows more strictly the principle of making UTRAN alone responsible for all radio resource management. This is underlined by the introduction of a generic and uniform control protocol for the Iu interface.

1.2.4 Bearer Architecture

As stated earlier in this chapter, the 3G system mainly acts as an infrastructure providing facilities, adequate bandwidth and quality for end-users and their applications. This facility provision, bandwidth allocation and connection quality together is commonly called Quality of Service (QoS). If we think of an end-to-end service between users, the used service sets its requirements concerning QoS and this requirement must be met everywhere in the network. The various parts of the UMTS network contribute to fulfilling the QoS requirements of the services in different ways.

To model this, the end-to-end service requirements have been divided into three entities: the local bearer service, the UMTS bearer service and the external bearer service. The local bearer service contains mechanisms on how the end-user service is mapped between the terminal equipment and Mobile Termination (MT). MT is the part of the UE that terminates radio transmission to and from the network and adapts the terminal equipment capabilities to those of radio transmission. The UMTS bearer service in turn contains mechanisms to allocate QoS over the UMTS/3G network consisting of UTRAN and CN. Since the UMTS network attaches itself to external network(s), end-user QoS requirements must be handled towards the other networks too. This is taken care of by the external bearer service.

Within the UMTS network, QoS handling is different in UTRAN and CN. From the CN point of view, UTRAN creates an "illusion" of a fixed bearer providing adequate QoS for the end-user service. This "illusion" is called the radio access bearer service. Within the CN, its own type of bearer service called the "CN bearer service" is used. This division between the radio access bearer and the CN bearer service is required since the QoS must be guaranteed in very different environments and both of these environments require their own mechanisms and protocols. For instance, the CN

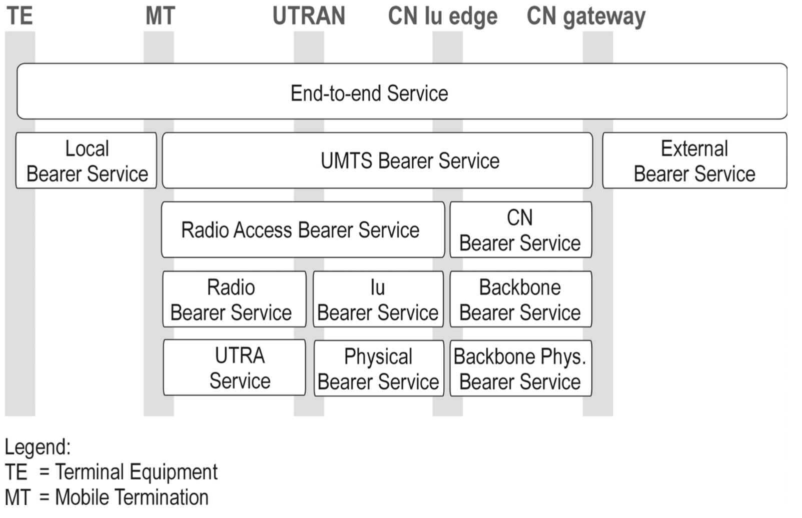

Figure 1.5 Bearer architecture in UMTS

bearer service is quite constant in nature since the backbone bearer service providing the physical connections is also stable. Within UTRAN the radio access bearer experiences more changes as a function of time and movement of the UE, and this sets different challenges for QoS. This division also pursues the main architectural principle of the UMTS network (i.e., independence of the entire network infrastructure from radio access technology).

The structure presented in Figure 1.5 is a network architecture model from the bearer and QoS point of view. Since QoS is one of the most important issues in UMTS, QoS and bearer concepts are handled throughout this book.

The rest of the book uses these architectural approaches as cornerstones when exploring UMTS networks and their implementations.