computer communic@tions and networks

UMTS

Origins, Architecture and the Standard

Springer

Pierre Lescuyer

Translation Editor: Frank Bott

Computer Communications and Networks

Springer-Verlag London Ltd.

The **Computer Communications and Networks** series is a range of textbooks, monographs and handbooks. It sets out to provide students, researchers and non-specialists alike with a sure grounding in current knowledge, together with comprehensible access to the latest developments in computer communications and networking.

Emphasis is placed on clear and explanatory styles that support a tutorial approach, so that even the most complex of topics is presented in a lucid and intelligible manner.

Also in this series:

An Information Security Handbook
John M.D. Hunter
1-85233-180-1

Multimedia Internet Broadcasting: Quality, Technology and Interface
Andy Sloane and Dave Lawrence (Eds)
1-85233-283-2

The Quintessential PIC Microcontroller
Sid Katzen
1-85233-309-X

Information Assurance: Surviving in the Information Environment
Andrew Blyth and Gerald L. Kovacich
1-85233-326-X

Pierre Lescuyer
Translated and edited by Frank Bott

UMTS: Origins, Architecture and the Standard

Springer

Pierre Lescuyer
Nortel Networks, Parc D'activité de Magny Chateaufort, 78928 Chateaufort, Yvelines Cedex 9, France

Translated and edited by
Frank Bott
Department of Computer Science, University of Wales, Penglais, Aberyswyth, SY23 1EY, UK

Series editor
Professor A.J. Sammes, BSc, MPhil, PhD, FBCS, CEng
CISM Group, Cranfield University, RMCS, Shrivenham, Swindon SN6 8LA, UK

British Library Cataloguing in Publication Data
Lescuyer, Pierre
UMTS : origins, architecture and the standard. – (Computer communications and networks)
1. Universal Mobile Telecommunications System
I. Title II. Bott, Frank
621.3′8456
ISBN 978-1-85233-676-9 ISBN 978-0-85729-420-3 (eBook)
DOI 10.1007/978-0-85729-420-3

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

Computer Communications and Networks ISSN 1617-7975
ISBN 978-1-85233-676-9

http://www.springer.co.uk

Translation from the French language edition of UMTS: Les origines, l'arhitecture, la norme
© Dunod, Paris 2002.

This work has been published with the help of the French Ministére de la Culture – Centre national du livre.

Typeset by Florence Production, Stoodleigh, Devon, England

34/3830-543210 Printed on acid-free paper SPIN 10887111

Preface

Since the beginning of the 1990s, cellular communication services have been enjoying an unprecedented level of development, made possible by the existence of the so-called second generation digital technologies, GSM (Global System for Mobile communication) being one of the most popular. These technologies, generally incompatible with each other, are the result of standards created at the start of the 1980s. After some years of evolution and successive improvements, these different standards have today reached the limits of their possibilities.

In order to permit the creation of new services and to offer users real mobility on a global scale, it has become necessary to make a technological jump and cross the threshold to third generation cellular networks. A number of partners (telecommunications equipment suppliers and operators) have therefore been working together for some years to define the future technology, trying to reconcile the requirements of new services (high quality wireless Internet, multimedia, etc.) with the necessity of guaranteeing to users and network operators the smoothest possible transition to the new generation.

Despite the unification efforts of the International Telecommunications Union, through its IMT-2000 program, there exist not one, but several third generation technologies, the main one being the Universal Mobile Telecommunication System (UMTS). Since January 1999, the 3GPP (Third Generation Partnership Project), in charge of the development of the UMTS standard, has carried out a substantial amount of work, whose concrete expression consists of several tens of thousands of pages of specifications spread over more than 300 documents.

The aim of this book is to present, from a system point of view, the architecture and the techniques employed in UMTS networks. The introductory chapters describe the origins of UMTS and its place among the third generation technologies. The succeeding, more technical chapters describe different aspects of UMTS, including the architecture, the structure of the radio interface, the protocols used, and the importance of the GSM inheritance.

The book is intended for students and telecommunications professionals who want a general view of UMTS, without losing themselves in the labyrinth of the specifications in the standard.

Contents

1 UMTS and Third Generation Networks

1.1 Origins of UMTS

For some years now, cellular telephony systems have been experiencing a level of growth unprecedented in the world of telecommunications. When the first cellular technologies were brought into service, at the beginning of the 1990s, we saw a rather slow take-off in the curve of subscriber numbers, hardly presaging the subsequent spectacular growth. At the beginning of 2001, there were some 450 million GSM subscribers in the world (including more than 40 million in the UK) (Figure 1.1). So-called second generation wireless technologies as a whole (GSM, IS-95, AMPS, PDC, PHS) serve a total of more than 600 million subscribers.

1.1.1 A Promising Future

Figure 1.2 shows global forecasts, in terms of subscribers, for radio-telephony technologies as a whole. They show that the present tendency, far from slowing down,

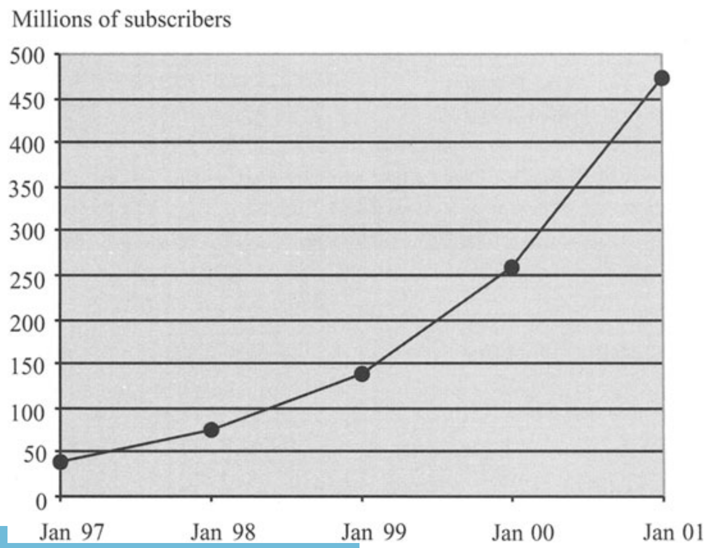


Figure 1.1 Growth in the number of GSM subscribers worldwide (source: *GSA*).

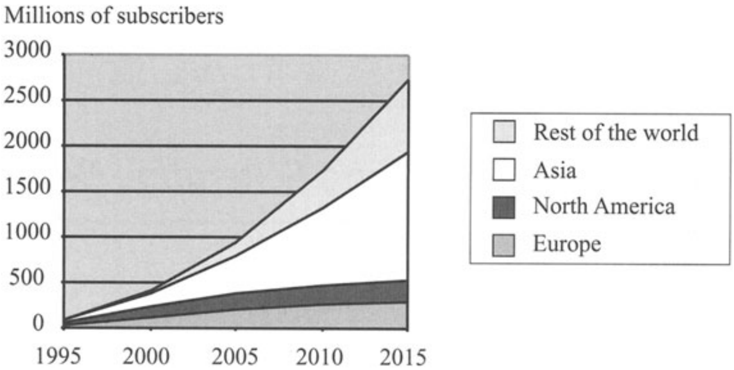


Figure 1.2 Growth forecast for wireless telephony (source: *UMTS Forum*).

is set to accelerate. It is even predicted that the number of subscribers to wireless services will overtake the number of subscribers to wired services by 2004.

These forecasts are based on the following observation: the European and North American markets, already very developed, will reach their maximum level when the level of equipment provision reaches 80% of the population. Applying this reasoning to Asia, to the Pacific, and to the rest of the world (South America, etc.), we realise that the development potential of these regions is far ahead of that of Europe and North America.

It could well be that these forecasts fall short of the reality. It is, in fact, perfectly possible that making new services, GPRS (General Packet Radio Service) for example, available within the GSM domain will accelerate still further the development of mobile telephony usage. Recent changes to the GSM standard will permit the use of the radio communications spectrum for packet-switched data services to be rationalised. This should open the way for such new applications as the use of telemetry for domestic meter reading or remote diagnostic applications, whether medical or mechanical. Furthermore, the systematic use of cellular telephony in the professional domain, linked to the need to separate private and professional usage of the mobile, may have a by no means negligible effect on the growth of the market. This separation will, in fact, lead users to subscribe twice, once for professional use and once for private use.

1.1.2 Incompatible Systems

There is still, however, one spectre at the feast. To meet the spectacular growth of recent years, some very different and, unfortunately, incompatible technologies have been used. Among the most widely used second generation systems, there are:

- PDC (Personal Digital Cellular) and PHS (Personal Handyphone System), used mostly in Japan;

Table 1.1 Principal technological differences between GSM, IS-95, IS-136 and PDC.

	GSM	IS-95	IS-136	PDC
Modulation method	GMSK	QPSK	DQPSK	QPSK
Access method	TDMA/FDMA	CDMA	TDMA	TDMA
Frequency band (MHz)	900/1 800/1 900	800/1 900	800/1 900	800/1 400
Bandwidth (kHz)	200	1,250	30	25
Usage	Worldwide	American continent, Asia	North America	Japan

- AMPS (Advanced Mobile Phone Service), IS-136 or D-AMPS (Digital-AMPS), the digital successor to AMPS, and IS-95 (also known under the commercial name of cdmaOne), used principally on the American continent and in Asia;
- GSM (Global System for Mobile communication), adopted in the majority of countries offering cellular communication services. GSM was originally a purely European standard. Where GSM has been deployed outside Europe (for example in North America or Asia), it is often in competition with another system.

The incompatibility between these systems manifests itself in several different ways.

First of all, major differences in the use of the radio segment prevent any mobile from working over a network that uses a technology different from that for which it was designed. The differences between the radio technologies used by second generation systems lie in such fundamental aspects as the modulation method, the frequency band used, and the structure of the physical channels. Table 1.1 illustrates some of these differences for the four main systems of cellular radio communication: GSM, IS-95, IS-136 and PDC.

It is, of course, always possible to design multimode terminals able to function as well on one of these technologies as on another: IS-95 and GSM, for example. Attempts to do this have rarely proved successful, however, because of the cost and the extra bulk of such equipment.

Secondly, subscriber management is not at all the same on the different systems. In the case of GSM, the subscriber has a SIM (Subscriber Identification Module) card, containing all the information about the subscriber (including, for example, the subscriber identity or IMSI). This card is designed to be used in any GSM mobile and allows subscribers to continue to use the services when they move to a GSM network other than the one provided by the operator with whom they have subscribed; this is described as *roaming*. Other systems exist, however, notably in North America, in which the information about the subscriber is physically linked to the mobile equipment, thus limiting severely the possibilities for roaming.

1.1.3 Towards a Common Standard

The incompatibility among second generation systems has many inconveniences.

Unfortunate travellers who go to countries where the technology offered by their operator is not represented find themselves suddenly deprived of their communication tool. This has a direct influence on the revenues of cellular operators. Operators who have chosen minority cellular technologies will suffer a loss of income because roaming subscribers cannot connect to their networks.

Developing a cellular communication technology constitutes a major effort for an equipment supplier, which has a direct effect on the selling price of the systems. The multiplication of cellular technologies prevents the operators from benefiting from the economies of scale that a common technology, deployed on a global scale, would bring.

Since the idea of a third generation radio communication system began to germinate, there has been a strong desire on the part of the operators to see a common standard defined.

1.2 The Place of UMTS Among Third Generation Networks

As already explained, the second generation radio-telephony scene is made up of a multitude of different technologies. For each of these systems, there is a standardisation body responsible for the specifications and for their development. In order to avoid a repeat of such a scenario, it became necessary to entrust the definition of the third generation to as independent and representative an organisation as possible. The International Telecommunications Union (ITU) was the obvious body to fulfil this role.

Thus it was that the ITU came to define the IMT-2000 concept, seeking to bring together the proposals of the different standardisation bodies and to define an international standard with the following objectives:

- support for multimedia applications;
- support of higher speeds (up to 2 Mbps);
- extended roaming, allowing the subscriber to benefit from a service coverage well beyond that available today.

The first IMT-2000 milestone (Figure 1.3) was the end of the technical proposal submission phase for the most sensitive link in the system: the radio segment, called Radio Transmission Technology (RTT) by the ITU. By June 1998, 16 proposals had been submitted, coming from 12 different standardisation bodies. They included

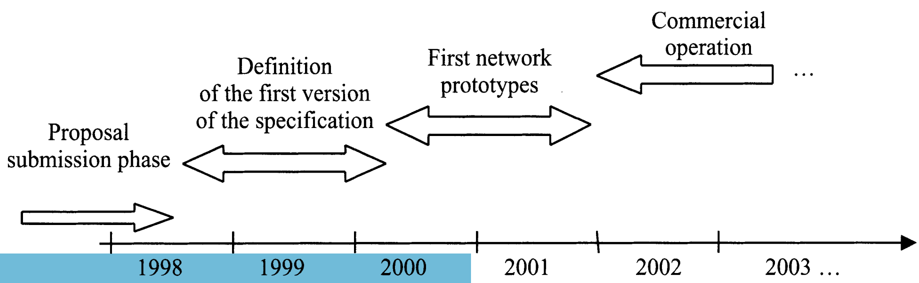


Figure 1.3 IMT-2000 timetable.

Table 1.2 The RTT proposals submitted to the ITU.

	Description	Type	Origin
UWC-136	Universal Wireless Communications	Terrestrial	TIA
WIMS	Wireless Multimedia and Messaging Services	Terrestrial	TIA
NA: WCDMA	North American wideband CDMA	Terrestrial	T1P1
cdma2000	Wideband CDMA (based on IS-95)	Terrestrial	TIA
SAT-CDMA	Network of satellites in low orbit	Satellite	TTA
DECT	Digital Enhanced Cordless Telecommunications	Terrestrial	ETSI
TD-SCDMA	Time-Division Synchronous CDMA	Terrestrial	CATT
SW-CDMA	Satellite Wideband CDMA	Satellite	ESA
SW-CTDMA	Satellite Wideband Hybrid CDMA/TDMA	Satellite	ESA
ICO RTT	Network of satellites in medium orbit	Satellite	ICO
W-CDMA	Wideband CDMA	Terrestrial	ARIB
CDMA II	Asynchronous DS-CDMA	Terrestrial	TTA
CDMA I	Multiband Synchronous DS-CDMA	Terrestrial	TTA
UTRA	Universal Terrestrial Radio Access	Terrestrial	ETSI
Horizons	Horizons Satellite System	Satellite	Inmarsat

proposals both for terrestrial communication networks and for satellite-based systems. They are briefly summarised in Table 1.2.

Figure 1.4 shows a breakdown of the proposals for the terrestrial network on the basis of the network access technology. As will be seen, CDMA is the dominant technology. This represents a break for the GSM operators group, more especially for the European operators. The move to CDMA will impose on them new methods of network deployment and new tools for cellular planning.

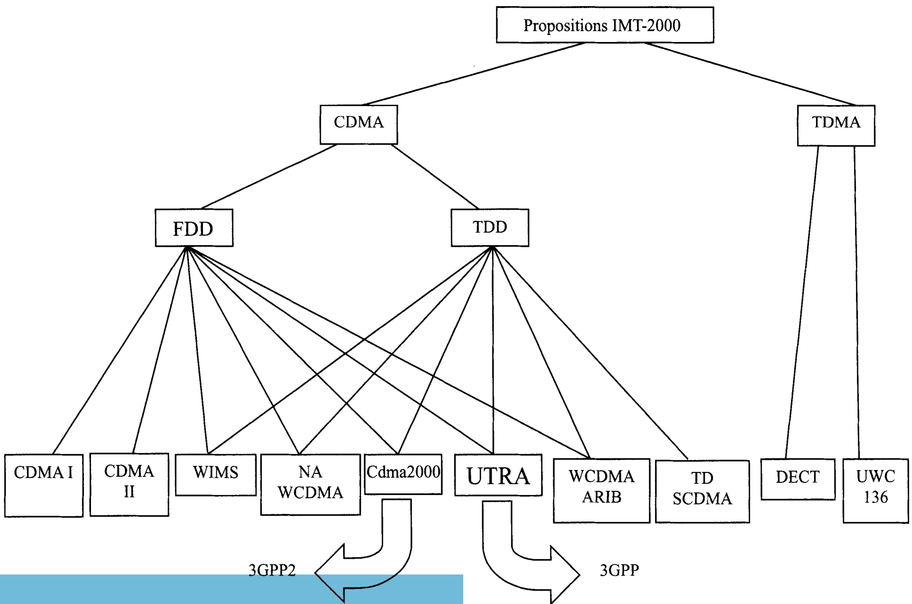


Figure 1.4 IMT-2000 proposals.

(The access network is the equipment associated with the radio segment of the network, which inter alia allocates radio resources and multiplexes users on to the radio channels. The UMTS access network is known by the name of UTRAN (Universal Terrestrial Radio Access Network).

The core network consists of those elements of the network responsible for managing subscribers, routing calls, and interconnection with external communications networks.)

The majority of the CDMA proposals include the two modes: TDD (Time Division Duplex) and FDD (Frequency Division Duplex). The FDD mode requires the use of a pair of frequency bands (the up and down paths use different frequency bands) while the TDD mode uses only a single band (the up and down paths are time division multiplexed). The proposals that use TDD have an advantage over the others: they are more flexible in the face of the problems of spectrum allocation encountered by the regulatory bodies. As we will see in the section on the IMT-2000 radio spectrum, this point is not without importance.

At the end of a negotiating phase, two technology families finally emerged from the set of proposals regarding the terrestrial networks, leading to the creation of two groupings of suppliers and network operators, whose principal features are summarised in Table 1.3:

- the Third Generation Partnership Project (3GPP), originators of the Universal Mobile Telephone System (UMTS), the specification of which inherits a number of features from the GSM standard;
- 3GPP2, created as a reaction to 3GPP's strong orientation towards GSM, in order to ensure the continuing existence of systems of the IS-95 type, the North American competitor to GSM.

These two groupings are interesting because they constitute a sort of 'neutral' environment, in which each participant, whether operator or supplier, from Europe,

Table 1.3 Principal features of 3GPP and 3GPP2.

	3GPP	3GPP2
Date of creation	January 1999	January 1999
Technology	UMTS	cdma2000
Affiliated bodies	ETSI (Europe)	TIA (USA)
	TTA (Korea)	TTA (Korea)
	TTC (Japan)	TTC (Japan)
	ARIB (Japan)	ARIB (Japan)
	T1 (USA)	CWTS (China)
Type of core network	MAP (GSM)	ANSI-41
Access network technology	DS-W-CDMA (FDD)	DS/MC-W-CDMA
	TD/CDMA (TDD)	(IS-95)

ARIB: Association of Radio Industries and Businesses (Japan)
CWTS: China Wireless Telecommunications Standard Group (China)
CATT: China Academy Telecommunications Technology (China)
ETSI: European Telecommunications Standards Institute (Europe)
TTA: Telecommunications Technology Association (Korea)
TTC: Telecommunications Technology Committee (Japan)
TIA: Telecommunications Industry Association (USA)

Asia or North America, has the same voting rights as the other participants. This is not the case with traditional standards organisations such as ETSI, T1 or ARIB, where only the local participants (for example, European operators and suppliers in the case of ETSI) have decision making powers.

In the 3GPP, decisions are taken on the basis of consensus. This means that, following a period of discussion and negotiation, every decision, before being adopted, must meet with the support of all the parties concerned. This way of working is fairly similar to that of many standardisation bodies.

Despite its initial objectives, IMT-2000 has not resulted in a unique standard. A consequence of this important point is the need to develop multimode terminals, the only way of ensuring continuity of service between networks supporting different technologies. However, the heritage of GSM in UMTS (in particular in the architecture of the core network) puts it in a favourable position. We will see, in fact, when we consider the objectives of UMTS, that continuity of service with GSM is one of its fundamental principles.

As Figure 1.5 shows, GSM has the benefit of a strong position. Its principal competitors, IS-95, PDC and IS-136, are clearly losing ground. The large number of GSM users is a factor that will play a major role in ensuring the future success of UMTS.

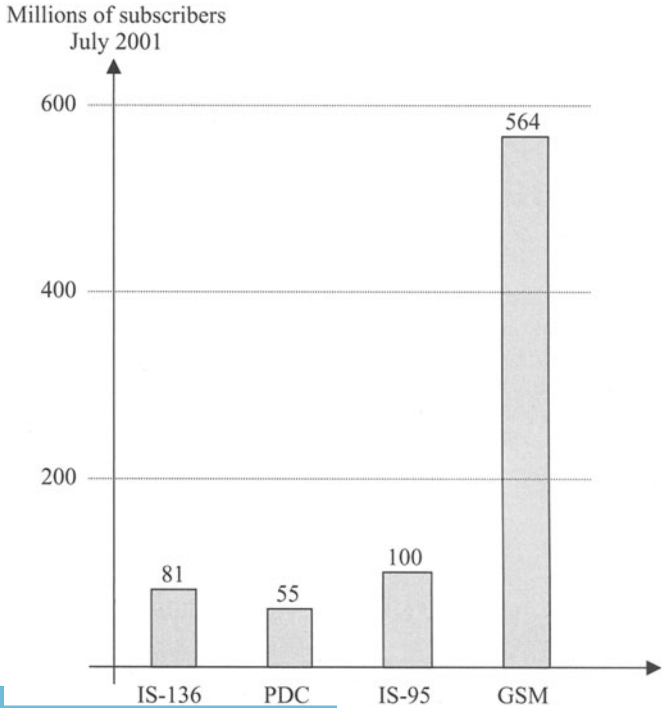


Figure 1.5 Distribution of mobile subscribers by technology (source: GSA).

Despite the large number of technologies proposed for IMT-2000, this strength means that UMTS is well placed to dominate the radio communication landscape of the third generation. Furthermore, thanks to the inclusion of Asian and North American partners, the creation of the 3GPP consortium has avoided UMTS becoming solely a European standard. If this had not been done, UMTS would have been excluded from some markets, just as happened to GSM.

1.3 Importance of Standardisation

The standardisation phase has been of particular importance within the UMTS programme for three reasons.

First, the 3GPP grouping includes many more supplier and operator participants than the standardisation bodies that defined the second generation standards. The options and technical choices decided on by 3GPP therefore carry much more weight.

Secondly, as we will see in later chapters, the process of specifying 3GPP has been much more thorough than that of defining the second generation standards. For example, standardised open interfaces are much more numerous in UMTS than in the GSM and IS-95 standards.

The final point concerns the value of patents in the world of the telecommunications industry. For the companies that build telecommunications equipment, the standard has become a real tool for protecting investment in research and development, and intellectual property more generally, all the more so in the case of a standard like UMTS that is intended to serve as a global reference standard.

1.4 Structure of 3GPP

The 3GPP grouping is made up of a number of separate technical specification groups (TSG), corresponding to distinct work areas (Figure 1.6). Each TSG is responsible for defining and producing specifications within its domain of interest. Initially, there were four TSGs:

- The purpose of the SA (Service and System Aspects) TSG is to specify the user services and the general architecture of the UMTS network. It also handles problems connected with security and confidentiality of communications carried on the network.
- The Core Network (CN) group is in charge of the call control protocols and of supplementary services, as well as interconnection to external networks.
- The Radio Access Network (RAN) group is responsible for the definition of the protocols and the architecture of the UMTS access network.
- The objective of the Terminals (*T*) group is to define the structure of the USIM card (the successor to the SIM card used in GSM mobiles), along with the functions and conformance tests for UMTS terminals.

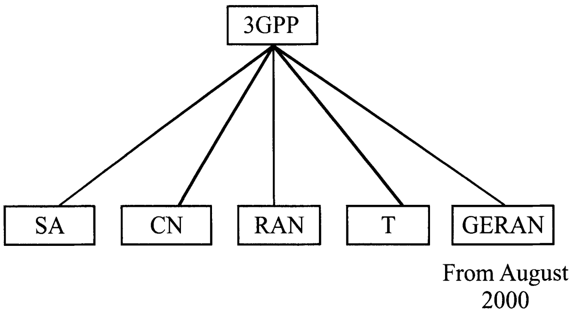


Figure 1.6 Structure of 3GPP.

In August 2000, a new TSG was added: GERAN (GPRS EDGE Radio Access Network) is responsible for the evolution of the access network part of the GSM standard. This new group results from the migration of part of these activities from ETSI to 3GPP. This development is a good illustration of the desire to define UMTS as a communication network that includes different access technologies, as we will see in Chapter 3.

1.5 Objectives of UMTS

1.5.1 Compatibility with Second Generation Systems

The deployment of second generation cellular systems has been, and still is, an extremely expensive business for the operators. The cost of a network with national coverage is measured in hundreds of millions of pounds, invested over several years. The investment becomes profitable only after a period of operation of at least five or six years.

As a result of their complexity and the growth forecasts for wireless telephony, the level of operators' investment in third generation networks is likely to be even higher. In consequence, the future third generation standard must provide for a certain level of compatibility with existing systems, in order to allow operators of second generation technology to extend the life of their investments and also to provide continuity of service to users during the migration to the new technology.

Compatibility between UMTS and GSM includes two different aspects:

- compatibility in terms of services offered to the user;
- transparency of the network, as seen by the user.

Compatibility of Services

In general, we can classify the services offered by a communication network into three groups:

- bearer services;
- tele-services;
- supplementary services.

Bearer Services

The term 'bearer service' refers to a transport service provided by a transmission channel set up between users. To each class of bearer service there corresponds a transmission capacity endowed with certain characteristics. In GSM, the bearer services are characterised by a list of attributes including, inter alia:

- speed (1200, 2400, 4800, 9600 bps);
- type of access (synchronous, asynchronous);
- transmission mode (transparent, non-transparent);
- configuration (symmetric, asymmetric).

Tele-services

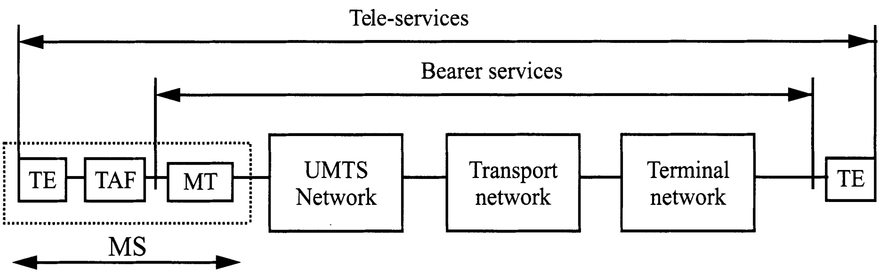
Tele-services include the principal services that the network offers to the user. In GSM they include:

- classical telephony;
- emergency calls;
- short messages initiated by the mobile, by the network, or broadcast;
- Group 3 fax.

Figure 1.7 illustrates the difference between bearer services and tele-services in the case of a link between a mobile and a fixed terminal. The bearer services include a transmission network made up of a number of sub-networks:

- the UMTS network (itself made up of a radio interface and a terrestrial network);
- the transit network (for example the PSTN, i.e. the public switched telephone network);
- and finally, a possible terminal network.

Tele-services as such are established from end to end, including adaptation to the bearer service provided by the network (TAF: Terminal Adaptation Function).



MS : Mobile Station
MT : Mobile Termination
TE : Terminal Equipment
TAF : Terminal Adaptation Function

Figure 1.7 Bearer services and tele-services.

Supplementary Services

Supplementary services bring together a whole range of services complementary to the tele-services. They are the subject of special management for the operator (and for the subscriber) because, in contrast to bearer services and tele-services, they can be invoiced independently and are activated and deactivated by the user.

A large number of supplementary services were defined within the GSM framework. Among the best known and the most frequently used in present day GSM networks are the following:

- caller identification (CLIP: *Call Line Identification Presentation*);
- different types of call forwarding (CFU: *Call Forwarding Unconditional*, CFNRy: *Call Forwarding on No Reply*, etc.)
- services connected with double calls (CW: *Call Wait* and HOLD: *Call Hold*).

From the very first phase of UMTS (version 99 of the 3GPP specifications), it was decided that a UMTS network should be capable of offering as a minimum the set of bearer services, tele-services and supplementary services defined within GSM. This constraint is of great importance, both for the users and for the operators of GSM networks, since it will allow continuity of service to be assured during the migration to third generation networks.

To complement the set of services common to GSM and UMTS, some services specific to UMTS, such as *Multicall*, have been defined.

Transparency of the Network, as Seen by the User

As a general rule, cellular telephone operators try, as far as possible, to conceal from users the complexity of the methods needed to provide the services offered. This attitude has been apparent several times in the course of the development of GSM networks. We can cite two examples of this in developments carried out by GSM operators with the intention of increasing the call capacity of their network: microcellular networks and dual band networks.

Microcellular networks are made up of small cells (or *microcells*) to complement a large but saturated cell (*macrocell*). The cellular network is thus made up of two levels of coverage, the micro level and the macro level. The operator's strategy is then to distribute the active mobiles judiciously among the different cells of the micro level, in order to relieve congestion on the macrocell initially used.

In the case of dual band networks, the operator increases the density of the network's radio coverage by adding cells using a frequency band different to that used generally by the network. For example, an operator using the 900 MHz band might have recourse to cells using the 1800 MHz band, in order to improve the radio coverage of the network.

The deployment of these two techniques for increasing the density of the network coverage required changes to the standard and functional developments of the GSM networks in use, as well as modifications to the operating parameters of the networks. From the user's point of view, all these changes took place completely unobserved (except, perhaps, for the impression of an improvement in the quality of the

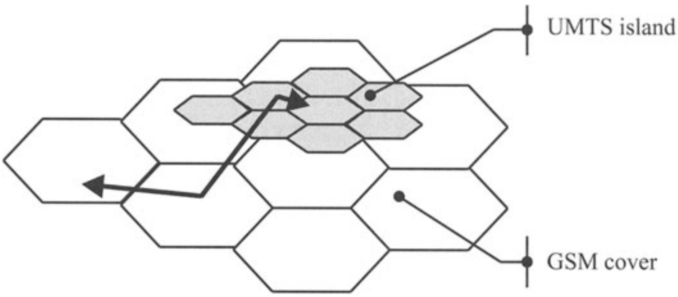


Figure 1.8 UMTS and GSM cohabiting.

service due to the increased capacity). The user passes from one layer of the network to another (or from one frequency band to another) without even realising it.

In the case of third generation networks, national coverage will certainly not be guaranteed at the start of the roll-out. In fact, since the GSM networks currently in use are not yet saturated, the operators have no particular reason for rushing ahead with the changeover to UMTS, apart from the desire to provide new services. On the other hand, to set up national coverage requires a very substantial investment, and the investments in the GSM network are still too recent to be written off.

For all these different reasons, the appearance of the first UMTS networks to be brought into use will doubtless be quite similar to that shown in Figure 1.8: broad GSM coverage completed by a few small UMTS islands placed in areas of dense traffic. It is therefore necessary for the operators, who will be providing UMTS and GSM services at the same time, to guarantee to their subscribers as smooth a transition as possible at the frontiers between the second and third generation technologies. When the user moves outside the area of UMTS coverage, it is important to avoid any sharp break of service while searching for the equivalent service in the GSM part of the network. Conversely, when a user with a multi-mode GSM/UTMS mobile is found to be in a zone covered by both technologies, the traffic management strategy can mean that the operator wants the user to use the UMTS resources. Such a traffic management policy would allow congestion on the GSM part of the network to be relieved and avoid the risk of saturation.

Arising from this, right from the beginning of the UMTS standard, a large number of functions intended to guarantee continuity of service between GSM and UMTS networks have been defined,.

1.5.2 Support for Multimedia

One of the primary objectives of IMT-2000 is support for multimedia applications.

Multimedia support is the capacity simultaneously to accept (for a terminal) or to deliver (for the network) services of different natures, for example, voice, videotelephony, file transfer or Web navigation. True multimedia support must be capable of offering a combination of several of these services simultaneously,

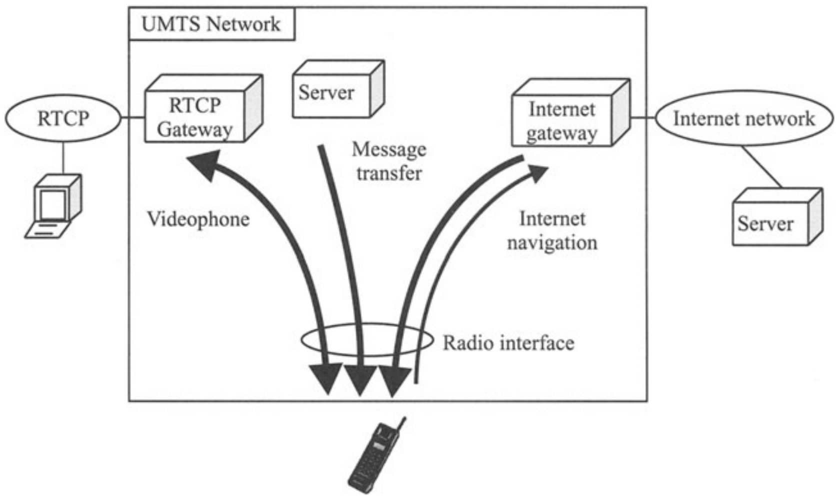


Figure 1.9 Multimedia networks.

while still being able to add or remove a service without disturbing those that the subscriber is actually using (Figure 1.9).

Multimedia support is not really a new idea in the world of cellular telephony. The GPRS standard offers embryonic multimedia support. Under the GPRS standard, class A mobiles are, in fact, able to offer classical telephony services or circuit mode data exchange, simultaneously with data applications in packet mode.

The objective of the third generation networks is to generalise this idea.

1.5.3 Transfer Rates Supported

The growing public demand for high performance communications has been increasingly evident in recent years. One can see this tendency in the provision of rapid Internet access technologies (such as cable or ADSL on fixed subscriber lines), allowing the user to obtain higher data rates than those offered by the modems currently available.

It is difficult to predict exactly what the flagship applications supported by the wireless networks of the future will be. In consequence, users' requirements in terms of performance remain an unknown. However, it is reasonable to expect that wireless networks will have to face a demand for increasing performance comparable to that experienced with fixed-line telephony.

As the successor to GSM, UMTS must offer a range of transfer rates beyond what is offered by the second generation. It was decided that UMTS would be designed in such a way as to guarantee the following transfer rates:

- 144 Kbps in an outdoor rural environment;
- 384 Kbps in an outdoor urban environment;

- 2 Mbps for short distance communication inside a closed building (i.e. reduced mobility).

Of course, the higher the performance required, the more severe the constraints on the radio environment. This is true even for the third generation technologies. Thus the maximum transfer rate of 2 Mbps will only be available in the best radio transmission conditions, that is to say, for nearly stationary users close to the broadcasting antenna of the cells in which they are located.

1.5.4 UMTS Service Classes

In order to cover the set of present and future service needs envisaged for UMTS, four service classes have been defined, grouping services together according to their respective constraints. The principal constraints taken into account in defining these UMTS service classes are:

- information transfer delay;
- variation in information transfer delay;
- tolerance to transmission errors.

Information transfer delay is particularly important for interactive applications with strong real-time constraints (for example, classical telephony or video-telephony). A deterioration, even slight (a few hundred milliseconds) in the transfer delay rapidly becomes unbearable for the user. (This is generally the case for satellite telephone connections.) On the other hand, it has little importance for Internet-based services (browsing, electronic commerce, etc.), for which users easily get used to a response time of the order of a second.

Variation in information transfer delay is equally critical for applications with real-time constraints where it is important that the gap between packets of information at the source is faithfully restored at the receiving terminal. In the case of classical telephony, specific constant-speed resources are allocated to the communication. As a result, the transfer delay suffers little variation. This is not the case with telephony applications or video broadcasting on the Internet, for which it is essential to adopt special mechanisms to compensate for the 'jitter' generated by the transmission network.

Tolerance to transmission errors is an important factor for data transmission applications. These applications (such as downloading of files, banking transactions or electronic commerce) require that the information is faithfully transmitted by the network. This is not the case with applications of the classical telephony type, which can accept a much higher error rate. Human perception is actually quite tolerant to transmission errors in telephone applications.

The four classes of service defined within the UMTS framework can be divided into two groups:

- class A (conversational) and class B (streaming) for application with real-time constraints;
- class C (interactive) and class D (background) for applications using data sensitive to transmission errors.

Class A: Conversational

This class includes all the bi-directional services involving two speakers, or even a group of people. The constraints associated with this class depend essentially on human perception. Thus, for these applications, the information transfer delays are limited to very low values (100 to 200 ms maximum). On the other hand, the tolerance of human perception to errors in the transmission of images or sounds means that the network can provide an acceptable quality of service even in the presence of transmission errors (Figure 1.10).

The services represented by class A include, among others, telephony, video-telephony, and interactive games.

Class B: Streaming

This class includes all the services involving a user and a data server. Class B has practically the same characteristics as class A, apart from a few differences:

- Applications are asymmetric, most of the data being transferred from the network to the mobile.
- Acceptable information transfer delays are larger. The absence of interaction between the user and the source allows this constraint, characteristic of class A applications, to be relaxed.

The information transfer delay can be long (in comparison with class A services) without the quality of service perceived by the user being affected, provided that the variation in this time remains limited (Figure 1.11).

The services represented by class B include, among others, video-on-demand, radio broadcasting, and image transfer applications.

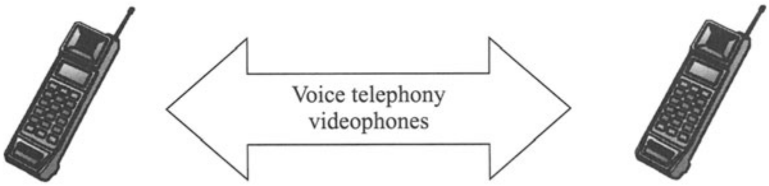


Figure 1.10 Class A.

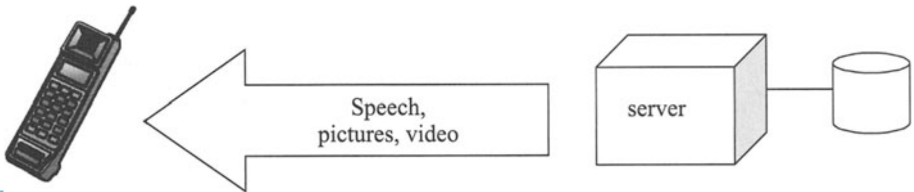


Figure 1.11 Class B.

Class C: Interactive

This class includes all the services in which a user sustains an interactive dialogue with an application server or a data server. Unlike classes A and B, class C does not require special real-time performance (except that the response to a user's request must still arrive within an acceptable time). On the other hand, it is essential for this type of application that the information transmitted does not suffer any alteration (Figure 1.12).

Services represented by class C include, among others, Internet browsing, file transfer using FTP, transfer of electronic messages, and all types of electronic commerce.

Class D: Background

The characteristics of class D are quite close to those of class C. The difference is that the information transmitted has lower priority than that of class C (Figure 1.13).

Class D applications include, amongst others, fax transfer, notification of electronic messages, and transmission of short text messages (SMS: *Short Message Service*).

Conclusion

Figure 1.14 shows the four service classes in terms of their respective characteristics.

The purpose of these service classes is to allow the UMTS network (in particular, the wireless part of the transmission path) to allocate resources and to protect

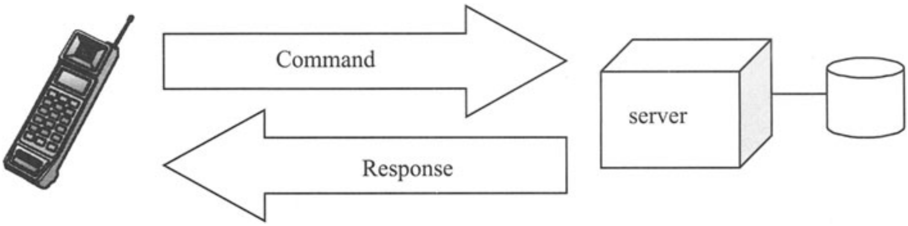


Figure 1.12 Class C.

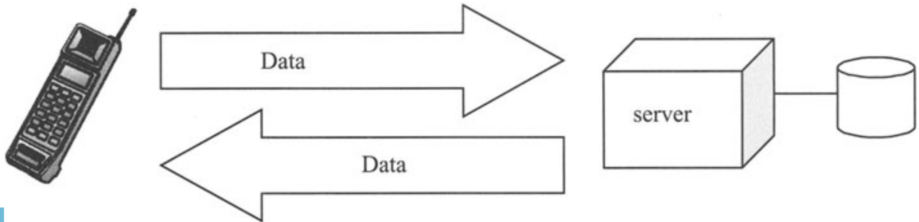


Figure 1.13 Class D.

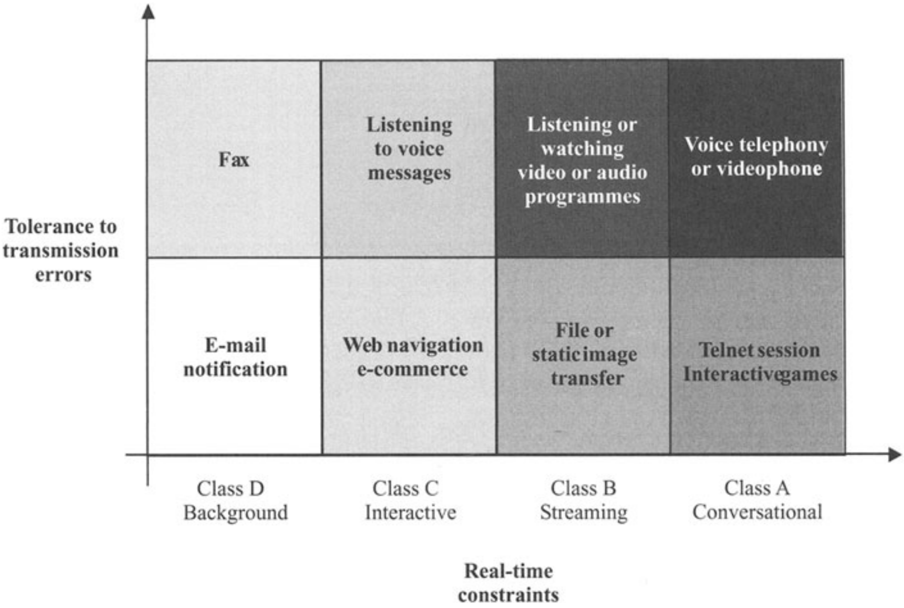


Figure 1.14 The service classes.

the data transmitted in a way appropriate to the service requested by the user. This is particularly important in the UMTS framework, where the UTRAN (the UMTS access network) has increased flexibility in comparison with second generation networks such as GSM.

As we will see later, the access network behaves like a service provider to the core network. Radio resources are allocated in accordance with the parameters of the requests made by the core network. The service class is one of the parameters.

1.6 Frequencies Allocated to the Third Generation

Following a study of future needs, the ITU decided, in 1992, to reserve 230 MHz of spectrum for IMT-2000, divided into two bands:

- 1885–2025 MHz for the first band;
- 2110–2200 MHz for the second band.

It also decided to reserve 150 MHz of spectrum, again divided into two bands, for satellite communication systems (MSS: *Mobile Satellite Service*).

The way the 230 MHz allocated to IMT-2000 were to be used was not specified by the ITU, leaving open the possibility of technologies using paired spectra or not.

At the ITU's World Radio Conference in June 2000, it was agreed that, in addition to the 230 MHz reserved initially, of the order of a further 160 MHz would be required by 2010. The conference identified three frequency bands that offered the greatest potential for meeting the predicted demand in a globally harmonised manner:

- 806–960 MHz;
- 1710–1885 MHz;
- 2500–2690 MHz.

Radio communication technologies using paired spectra are said to be of type FDD (Frequency Division Duplex). Uplink communication, that is from the mobile to the network, uses one band of frequencies, while downlink communication, from the network to the mobile, uses a different and disjoint band. The difference between the two bands is called the *duplex gap*. The GSM system uses this method of communication.

When, in contrast, a single frequency band is used, the uplink and downlink directions are multiplexed in time. We then speak of the TDD (Time Division Duplex) mode. This type of technology is used, for example, by DECT systems.

Figure 1.15 shows how the frequencies used today by certain second generation cellular communication systems are positioned within the frequency band initially allocated to IMT-2000:

- DECT;
- DCS1800 (GSM technology used in the 1800 MHz);
- PCS1900 (GSM technology used in North America in the 1900 MHz);
- PHS (Personal Handyphone System, used in Japan).

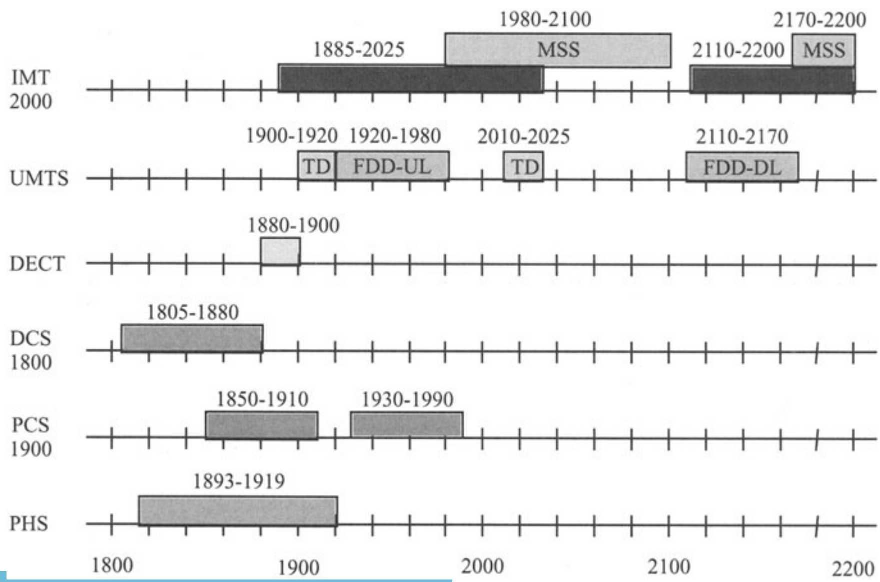


Figure 1.15 The spectrum reserved for IMT-2000.

The figure also shows the frequencies used by UMTS in the IMT-2000 spectrum. The UMTS frequencies are divided between the TDD and FDD technologies defined for the UTRAN, the UMTS access network. Two slots have been reserved for TDD, giving 35 MHz of bandwidth. FDD occupies 120 MHz of paired spectrum, i.e. 60 MHz for upward transmission (FDD-UL) and 60 MHz for downward transmission (FDD-DL).

It is apparent from this schema that the spectrum reserved for IMT-2000 usage is inevitably not accessible in all countries, because of the second generation technologies already in use. Because of this, the UMTS standard includes the possibility of using the radio resources in either TDD or FDD mode, in the hope of offering the maximum flexibility in the use of the resources of the spectrum.

1.7 Which Services Will UMTS Offer?

In view 2of the substantial sums that will be invested in setting up UMTS networks and bringing them into use, it is reasonable to wonder about the prospects for the services that the third generation networks will offer and for the revenues that they will generate. Figure 1.16 shows the results of a study on this subject published in the *UMTS Forum*.

It appears that classical telephony will continue to contribute a large proportion (more than 25%) of the income from cellular telephony in 2010. In contrast to classical telephony, third generation networks will derive a comparatively small proportion (6%) of traffic from advanced telephony services such as video-telephony or multimedia conferencing.

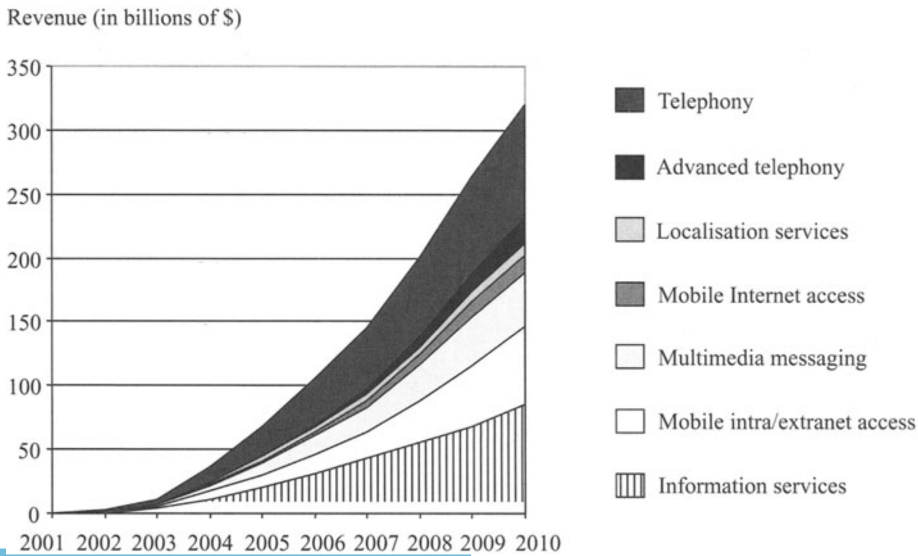


Figure 1.16 Forecasts of global revenue from 3G services (source: *UMTS Forum*).

In parallel with telephony services, we can see a clear increase in the proportion of the revenue linked to data services, especially:

- multimedia messaging, an extension of the GSM short message service, including one or more images, audio clips, video clips or animations;
- public Internet access through the services of an access provider or access to professional business networks;
- information services in the form, for example, of information portals or content servers, possibly personalisable by the user.

Almost completely absent from second generation networks in 2001 (the short message service apart), data applications will represent more than 60% of the income from third generation cellular telephony by 2010.

2 Review of Second Generation (GSM) Systems

2.1 Origins of UMTS

In the previous chapter, we saw that two large families had finally emerged from the set of proposals for the terrestrial access network submitted to IMT-2000: UMTS and CDMA2000, developed respectively by 3GPP and 3GPP2.

We also observed that these two technologies were developed in such a way as to maintain a certain continuity with the second generation networks currently in service. Thus, the UMTS standard inherited some concepts and architectural elements from the GSM standard, some elements of the GSM standard having being taken across without change while others were to a greater or lesser extent modified or improved.

In this chapter we will review some of the GSM concepts in order to understand better the contribution made by UMTS and to allow it to be seen in the context of GSM.

2.2 GSM Services

The first objective of the GSM standard was above all the provision of a telephone service, of a service for sending and receiving short messages (the SMS: *Short Message Services*), and of low speed, circuit-switched data transmission services (up to 9.6 Kbps). The standard later evolved to include higher performance data transmission services, both circuit switched (HSCSD: *High Speed Circuit Switched Data*) and packet switched (GPRS: *General Packet Radio Service*). These new techniques introduced into the GSM standard theoretically allow a user to benefit from data rates up to 171 Kbps (Figure 2.1).

The latest development of GSM (EDGE: Enhanced Data Rates for GSM Evolution) should even offer speeds up to a theoretical maximum of 345.6 Kbps in the most favourable cases (i.e. for a user using the full capacity of a GSM carrier, moving about slowly, in good radio propagation conditions).

2.2.1 High Speed Circuit Switched Data

HSCSD is an improvement to the circuit-switched data services of phase 2 of the GSM standard, intended to offer higher user performance than the 9.6 Kbps base service.

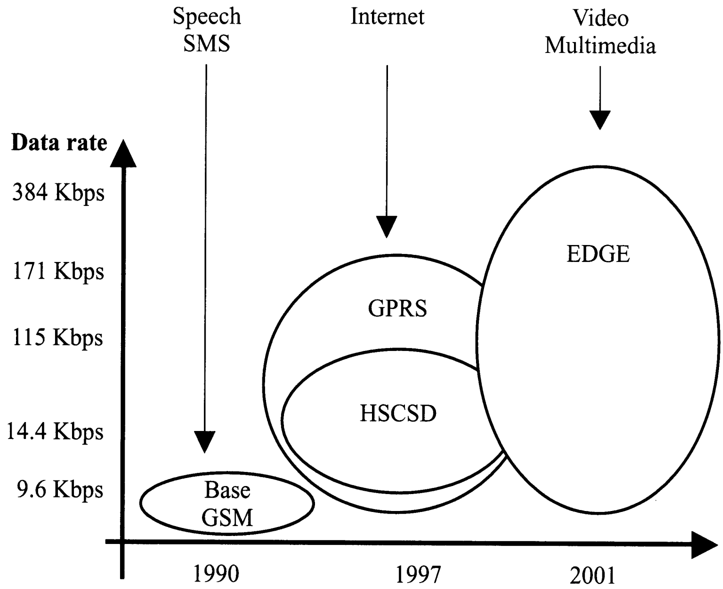


Figure 2.1 Development of GSM services.

In GSM, each carrier is divided into eight time intervals or *time slots* representing that many elementary resources. Each of these elementary resources can offer a maximum user performance of 14.4 Kbps, thanks to the use of special channel coding. Traditionally, each terminal uses just one of these elementary resources in the case of a classical telephone call or low speed data transfer. HSCSD allows up to eight of these resources to be combined for a given transmission, which allows a theoretical maximum of 8×14.4 Kbps, that is 115.2 Kbps to be attained.

In the case of data services in non-transparent mode, moreover, it is possible to vary the number of elementary resources used in the course of the transmission. This point is crucial in a cell change, in the course of which it is sometimes difficult to guarantee that the new cell will be able to offer the same number of elementary resources as the old one.

Unfortunately, this process of increasing or decreasing the number of elementary resources used is, in fact, quite slow and unresponsive, because it is based on the traditional GSM mechanisms for allocating and releasing resources. This feature makes HSCSD not very suitable for a number of today's Internet applications, which have fairly low mean data rates but require very high peak performance.

2.2.2 General Packet Radio Service

GPRS provides a general way of ameliorating the inefficiency of the allocation mechanisms for fixed capacity circuits imposed by HSCSD and the GSM data services, for example, for Internet-based applications.

The basic principle of GPRS is to share a certain number of the elementary resources of a cell among several users, by means of 'on-demand' allocation mechanisms, which are faster than the conventional mechanisms of the GSM standard. We can thus expect GPRS to give a more efficient utilisation of radio resources for packet-switched applications by avoiding allocating fixed resources to each user.

Like HSCSD, GPRS can allocate to a user, at any given instant, eight time slots on a GSM carrier, which allows it to achieve a peak rate of 8 × 21.4 Kbps, i.e. 171.2 Kbps, by using a channel coding method defined specifically for this service (CS-4).

2.2.3 Enhanced Data Rates for GSM Evolution

EDGE is an enhancement of GSM systems offering increased user data rates, thanks to a new modulation technique and new channel coding algorithms on the radio interface. The EDGE standard offers data rates up to 43.2 Kbps per GSM time slot.

2.2.4 Limits of GSM

These major contributions to the GSM standard, which can reasonably be expected to be the last, lie on the frontier between the technology of the present second generation networks and the future networks of the third generation. For this reason, we talk of 2.5G services.

Because the architecture of GSM systems was defined at the end of the 1980s, some significant compromises were necessary in order to maintain a certain compatibility with the networks then in use. In consequence, there are some serious limitations in the effective level of service provided to the user.

It might be thought that 2.5G services would be capable of competing with third generation networks so far as the services supported are concerned. It must, however, be understood that the maximum data rates indicated above are very theoretical. Amongst other things, they assume the use of a mobile device that is capable of using the eight time slots of a GSM carrier, which poses serious technical problems and is certainly not conceivable within reasonable limits of cost and bulk. Further, despite the relatively high rates offered in packet mode, GPRS offers no guarantee of real-time service, because of limitations connected with the GSM architecture. (This point is addressed in more detail in Chapter 8.)

It must, however, be recognised that the new services provided by GPRS represent a substantial break with the phase 1 and phase 2 GSM services, thereby constituting a first stage towards the third generation networks.

2.3 The Subscriber and the GSM Terminal

In this section we present a number of concepts relating to the subscriber and the GSM terminal. For the most part, these concepts have been retained in the UMTS standard.

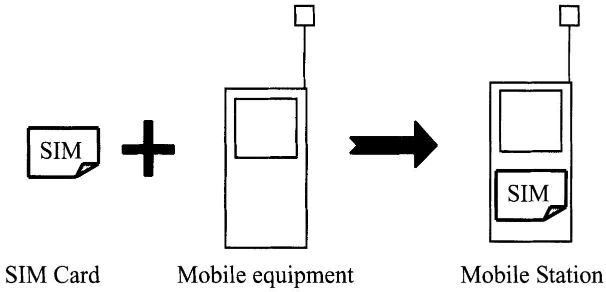


Figure 2.2 The terminal and the SIM card.

2.3.1 The Subscriber and the SIM Card

In GSM, as in many other second generation cellular systems, information about the identity of the subscriber is separated from the terminal equipment. This data is held on a smart card called a SIM (Subscriber Identification Module) card, which can be inserted into any piece of GSM equipment (Figure 2.2).

This system offers the subscriber the advantage of being able to change terminal equipment in a very flexible and completely autonomous way. Given the technological progress made by the manufacturers of mobiles, resulting in ever higher performance terminals being put on the market, the attractiveness of such a system is quickly apparent.

The primary function of the SIM card is to identify and authenticate the subscriber to the network. This requires certain data (such as the subscriber's authentication key or the IMSI, described below), which are by definition not modifiable and are recorded in a read-only way in the SIM card.

The SIM card also contains:

- temporary information that is more or less regularly modified by the network during the life of the SIM card, for example, the TMSI (Temporary Mobile Station Identity), or the current location area;
- information connected with the services to which the user subscribes.

The terminology used in the GSM standards distinguishes between terminals equipped with a SIM card and those without. When a SIM card is loaded into the terminal, it is referred to as a *mobile station*. The network is in a position to provide the service requested by the subscriber to the extent specified by the information stored in the SIM card.

When there is no SIM card in the terminal, it is referred to as mobile equipment. In this case, the only service request that the network can accept from the user of the equipment is an emergency call.

2.3.2 Subscriber Identification

The GSM subscriber is known to the network by a unique number called the *International Mobile Station Identity* (IMSI) (Figure 2.3).

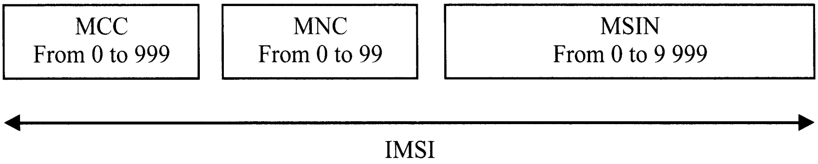


Figure 2.3 Structure of the IMSI.

The IMSI is made up of three parts:

- the *Mobile Country Code* (MCC), identifying the subscriber's home country;
- the *Mobile Network Code* (MNC), identifying the network operator (for example, Orange or Vodaphone);
- the *Mobile Station Identification Number* (MSIN) is the network operator's number for the subscriber.

To counter any attempt at fraud, the IMSI must remain secret so far as possible. This confidentiality is assured by a temporary identification mechanism intended to replace the IMSI during the call set-up procedure. Two types of temporary identifiers exist: the *Temporary Mobile Station Identifier* (TMSI), for the circuit-switched GSM services, and the *Packet TMSI* (P-TMSI), for the GPRS. The temporary identity is assigned by the network when the subscriber registers with the network, i.e. when the mobile is powered up. To strengthen this protection, a new TMSI or P-TMSI is allocated by the network either on a regular basis or when the mobile's location area is changed.

The mapping between the IMSI and the temporary identities allocated to the mobile is carried out by the MSC-VLR (see Section 2.4.1 below) for the circuit-switched GSM services and by the SGSN in the case of GPRS.

2.3.3 GSM Terminals

Terminal Identification

In the GSM standard, all terminals are identified by a unique number, in principle unfalsifiable, called the *International Mobile station Equipment Identity* (IMEI) (Figure 2.4).

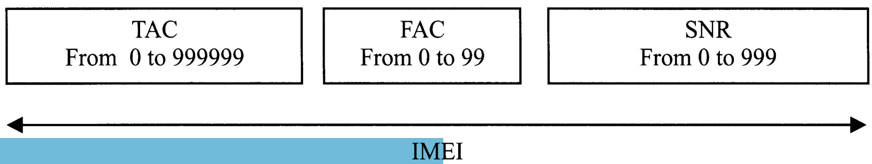


Figure 2.4 Structure of the IMEI.

The IMEI is made up of the following elements:

- the *Type Approval Code* (TAC);
- the *Final Assembly Code* (FAC, the assembly plant code);
- the *Serial Number* (SNR, serial number of the piece of equipment).

The value of the FAC and SNR fields is under the control of the manufacturer.

The IMEI is used optionally by GSM operators to fight thefts of terminals or to prevent terminals that behave in an incorrect manner or do not conform to the specifications from accessing the network. To this end, the operator maintains a database of 'forbidden' terminals, stored in a network node called the *Equipment Identity Register* (EIR). When the mobile establishes a connection with the network, for example, for the initial registration following power up or for setting up a call, the network can ask the terminal for its IMEI, and can as a consequence refuse access to a mobile identified in the EIR as suspect or stolen.

Terminal Classes

The GSM standard offers a large number of implementation options to the manufacturers of mobiles. The different options can be classified into two major groups:

- options connected with the hardware;
- options for the implementation of the different GSM services.

It is important that the network knows about certain of these options in order to avoid, for example, sending a text message to a mobile that is not capable of handling it. The list of terminal options communicated to the network is called the *class mark*. This information is sent by the mobile at various times, for example, when setting up a call or at the explicit request of the network.

Among the terminal options connected with the hardware are the following:

- the terminal's power classification, i.e. its maximum transmission power;
- the frequency bands supported: GSM, extended GSM, DCS1800, etc;
- multi-slot capacity, i.e. whether it can use several consecutive time slots, either in the uplink or downlink direction;
- support for EDGE modulation.

Options connected with the implementation of services include:

- encryption algorithms supported;
- the capacity to receive text messages;
- support for HSCSD;
- the class of GPRS service supported (A, B or C).

This last option is important because it constitutes a sort of gateway between GSM and UMTS at the service level.

GPRS defines three classes of service for the terminal, class A being the most complete. A terminal in class A in fact offers its user the possibility of being active simultaneously in GSM circuit-switched mode and in GPRS, i.e. the user can exchange data in GSM circuit-switched mode and in GPRS at the same time from a single terminal. One can thus imagine that future owners of a class A terminal will have the possibility, for example, of holding a telephone conversation at the same time as surfing the Internet or downloading information.

This class of services constitutes the embryo of the multimedia service that will be more extensively developed within the framework of UMTS.

2.4 GSM Network

2.4.1 Network Architecture

In very general terms, the GSM network is divided into two parts: the core network and the access network (Figure 2.5). In GSM terminology, the terms BSS (*Base station Sub-System*) and NSS (*Network Sub-System*) are more commonly used to describe these two parts of the network.

Access Network

The access network is that part of the whole network that manages the interface and the resources allocated on the radio interface. It also has an important role in managing the mobility of the user. Because the radio coverage is made up of cells of varying size (from a few tens of metres to a few tens of kilometres), the access network must be capable of passing the user from one cell to another during the course of a call. This is the *handover* function.

Core Network

The core network is the part of the network that manages the subscribers as a whole and the services provided to them, such as tele-services and supplementary services. The core network is responsible for call set-up and looks after the links between the GSM network and external networks.

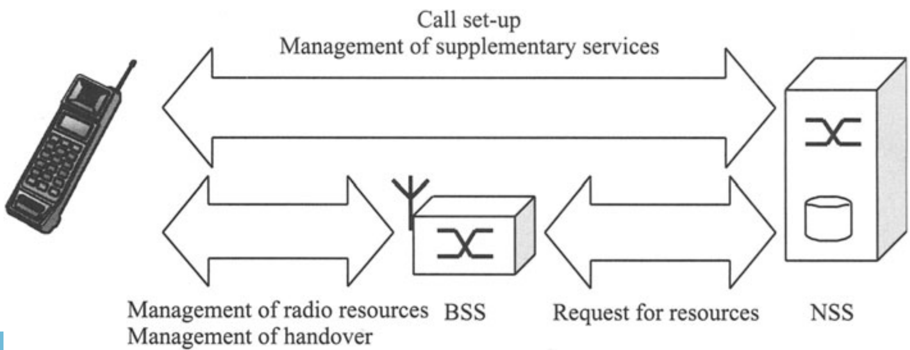


Figure 2.5 Overview of GSM network.

The access network also tries to make the radio technology used for access transparent to the core network. Ideally, the access network will behave towards the core network like a service provider, providing transmission channels in response to its requests. As we will see later, however, this principle of independence between the two sub-systems of the GSM network has not been completely respected.

Figure 2.6 presents a more detailed view of the architecture of the GSM network and of the internal and external interfaces. The network nodes in this diagram are in fact functional blocks that are not necessarily independent from a physical point of view. Indeed, it is completely possible to envisage a GSM network architecture complying with the standard in which, for example, the MSC and the VLR would be a single physical piece of equipment, or a BSC and a BTS that would be co-located in the same physical piece of equipment.

This figure includes both the network elements for circuit switching services (telephony, circuit-switched data transmission) and those for packet switching, used, for example, for applications accessible over the Internet.

The purely GPRS part of the core network (the SGSN and the CGSN) is made up of elements completely distinct from those used for the circuit-switching part. The only functions common to the two parts are the HLR, EIR and AuC functions.

Because GPRS was introduced late into the GSM standard, separating the circuit-switched applications from the packet-switched ones in the core network gives the network operators a certain flexibility during the migration to GPRS. The impact

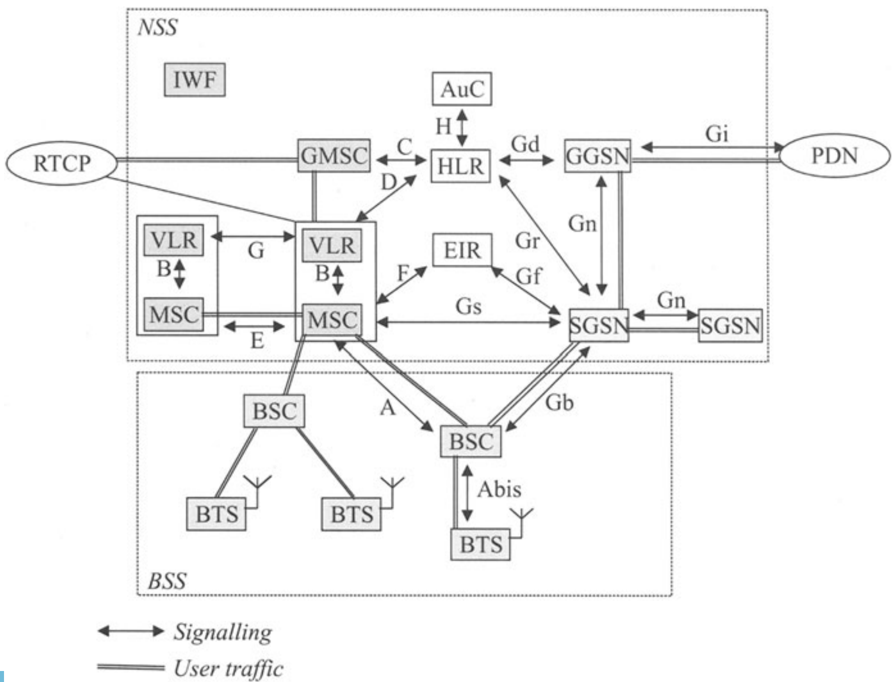


Figure 2.6 The GSM network.

of the deployment of GPRS on the equipment already in place is limited and does not therefore cause a break in the service offered to network users.

This separation of functions poses some problems, however. The proliferation of equipment makes the network operator's job of managing the network more complex. The duplication of certain functions in the core network (management of mobility, for example) also complicates the network configuration.

There exists a certain symmetry between the circuit-switching elements and the packet-switching elements of the GSM core network:

- The GMSC and CGSN each play a bridging role towards external networks (the Internet for CGSN, the Public Switched Telephone Network for the GMSC).
- The MSC/VLR and the SGSN each maintain information on the location of the subscriber, in circuit-switched and packet-switched modes respectively.

2.4.2 Constituents of the GSM Core Network

This section gives a quick description of the roles and functions of the constituents of the core GSM network. The implementation of the functions of the different nodes of the network is addressed in a more concrete way in Chapters 7 and 8.

Mobile Services Switching Centre (MSC) and Gateway MSC (GMSC)

The MSC is a data and signalling switch. It is responsible for managing the setting-up of calls with the mobile.

The GMSC is a special MSC that acts as a gateway between the GSM network and the PSTN (Public Switched Telephone Network). When someone wishes to reach a GSM subscriber starting from a point outside the GSM network (this is referred to as an incoming call), the call passes through the GMSC, which interrogates the HLR (see below) before routing the call to the MSC to which the subscriber is attached.

Home Location Register (HLR)

The HLR is the database containing information about the subscribers managed by the operator. For each subscriber, the HLR maintains the following information:

- account information (data services or not, registered for supplementary services, maximum data rate authorised, etc.);
- the identity of the mobile, i.e. the IMSI;
- the subscriber's calling number, or *Mobile Station International ISDN Number* (MSISDN).

The HLR also records the number of the VLR on which the subscriber is registered, in order to be able to reach the subscriber easily in the event of a call intended for him.

Visitor Location Register (VLR)

The VLR is a database attached to one or more MSCs. It is used to record the subscribers who are in a given geographical area called the *Location Area* (LA).

When the mobile moves and detects a change of location area, it must signal this change to the VLR. The network needs this update in order to be able to reach the mobile in the case of an incoming call.

The VLR holds much the same information as the HLR but, in addition, it keeps the following information for each subscriber:

- the temporary identity (TMSI) of the mobile, used to limit fraud linked to call interception and fraudulent use of the IMSI;
- the current location area of the subscriber.

In most networks, the MSC and the VLR are one and the same piece of equipment.

Authentication Centre (AuC)

The AuC is a network component that allows the network to carry out certain GSM security functions:

- authentication of the subscriber's IMSI;
- encryption of the call.

These two security functions are activated at the start of the call set-up with the subscriber. If either of these procedures fails, the call is rejected.

The AuC is coupled to the HLR and contains, for each subscriber, an identification key that allows it to carry out the authentication and encryption functions.

Equipment Identity Register (EIR)

The EIR is an optional component in GSM networks, intended to combat the theft of mobile terminals. The EIR is in fact a database containing a list, called the black list, of mobiles that are not allowed to access the network.

During call set-up, the network asks the terminal for its identity, i.e. its IMEI. If the IMEI returned by the terminal appears in the list of forbidden mobiles, the call cannot be established.

Of course, to be totally effective, this function requires that all subscribers who lose their terminals tell their operator of the loss or theft of the equipment. Equally, it is necessary for operators to update the database of their EIRs and to activate the identification procedures in their networks.

Interworking function (IWF)

The IWF serves as a gateway between the GSM network and the data networks (ISDN, PSDN, PDN). It converts between the data protocols used in the GSM network and those used in the data networks. IWF supports the *Radio Link Protocol* (RLP) used in GSM networks for the end-to-end transport of data in non-transparent mode.

Serving GPRS Support Node (SGSN)

The SGSN plays the same role vis-à-vis the GPRS part of the network as the VLR does for the circuit-switched part, i.e. it maintains a record of the subscriber's location, in the form of a *routing area* (RA), the RA being the GPRS equivalent of the location area. Like the VLR, the SGSN allocates a temporary identity to the mobile device, the *Packet-TMSI* (P-TMSI).

Gateway GPRS Support Node (GGSN)

The GGSN performs the same function for the GPRS part of the network as does the GMSC for the circuit-switched part, acting as a gateway to the external packet-switched networks (the Internet, private intranets, etc.).

2.4.3 Components of the GSM Access Network

The GSM access network consists of only two types of equipment: *base station controllers* (BSC) and *base transceiver stations* (BTS).

Base Station Controllers

A BSC is a piece of equipment capable of looking after one or more BTSs. The main functions of the BSC are the following:

- call routing between the BTS and the MSC;
- allocation of resources used on the radio interface. This function includes the initial allocation as well as control of the radio resources during the handover procedure;
- control of the BTS (start-up, supervision, activation of radio resources used by the call).

Base Transceiver Stations

The BTSs are the radio transmission equipment of the GSM network. They carry out various operations, including:

- coding and decoding of data transmitted over the radio interface;
- modulation and demodulation.

2.4.4 GSM Network Interfaces

In this section we briefly describe the different interfaces of the GSM network as well as their main functions (Table 2.1). A certain number of these interfaces have been standardised by ETSI, allowing operators to call on different suppliers for the components of their networks.

The following are the principal standardised GSM interfaces:

- the radio interface;
- the interfaces between the access network and the core GSM circuit-switched network and packet-switched networks;
- the MAP-E and Gn interfaces.

Table 2.1 GSM network interfaces.

Interface	Equipment	Main function
B	MSC-VLR	Exchange of user information and update of location area. This interface is not standardised because the functions of the MSC and the VLR are often combined in a single piece of equipment.
C	GMSC-HLR	Interrogation of the HLR in order to reach a mobile subscriber.
D	VLR-HLR	The VLR tells the HLR where the mobile is. The HLR gives the VLR information about the subscriber.
E	MSC-MSC	Handover management.
E	MSC-GMSC	SMS transport.
F	MSC-EIR	Verification of terminal identity.
G	VLR-VLR	Managing changes of location area.
H	HLR-AuC	Exchange of information necessary for encryption and authentication. This interface is not standardised.
Gb	SGSN-BSS	Transport of GPRS data between the core network and the access network.
Gc	GGSN-HLR	Exchange of information about the subscriber.
Gf	SGSN-EIR	Verification of terminal identity.
Gn	SGSN-SGSN	Management of GPRS mobility.
Gi	GGSN-PDN	Transport of GPRS data between the core network and the access network.
Gp	SGSN-GGSN	Management of GPRS mobility between operators.
Gr	SGSN-HLR	The SGSN tells the HLR the location of the mobile. The HLR provides the SGSN with information about the subscriber.
Gs	SGSN-MSC/VLR	Exchange of information about the location of the mobile, between the circuit-switched and packet-switched domains.
A-DTAP	MSC-BSC	Call set-up and close-down.
A-BSSMAP	MSC-BSC	Resource allocation and handover management.
Abis-O&M	BSC-BTS	Supervision of the BTS. This interface is not standardised.
Abis-RSL	BSC-BTS	Activation and de-activation of radio resources. This interface is not standardised.
Um	BTS-Mobile	Radio interface.

The Abis interface to the access network is not standardised, forcing the operator to use BSC and BTS equipment from the same manufacturer. On the other hand, thanks to the compatibility of the A and Gb interfaces, an operator can construct a network made up of BSSs (BSC + BTS) from different manufacturers. Figure 2.7 shows the possible combinations.

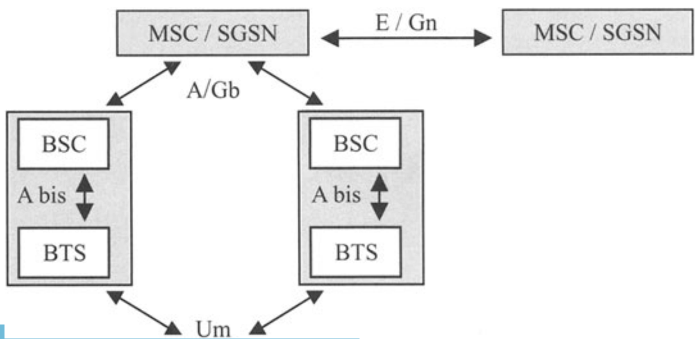


Figure 2.7 Supplier combinations in the GSM network.

2.4.5 GSM Radio Interface Protocols

This section gives a general view of the GSM network signalling protocols (Figure 2.8). These protocol layers do not completely conform to the layers of the OSI seven-layer model. One can, however, detect a certain equivalence for the low-level layers:

- level 1, or the physical layer;
- level 2, or the data link layer;
- level 3, or the network layer.

Level 1, based on GMSK modulation (or 8-PSK for EDGE), is a mixed TDMA/FDMA access mode (see Chapter 5). The specific GSM standard for level 1 of the radio interface is not covered in detail in this book.

The purpose of *level 2* is to provide reliable data transmission between two pieces of network equipment (i.e. for the radio interface, between the mobile and the network).

The level 2 protocol layers defined in the GSM standard for the circuit-switched and packet-switched parts are radically different. In the case of circuit-switched GSM, level 2 is based on the LAPDm protocol, which is itself a version of the LAPD protocol adapted to the radio interface. This adaptation, specific to the GSM standard, has not been retained in the UMTS standard and for this reason will not be considered further in this book.

The GPRS level 2 is much more complex than the LAPDm used in circuit-switched mode, because it offers the possibility of sharing one or more GSM radio resources (i.e. one or more time slots of a single TDMA GSM carrier) among several

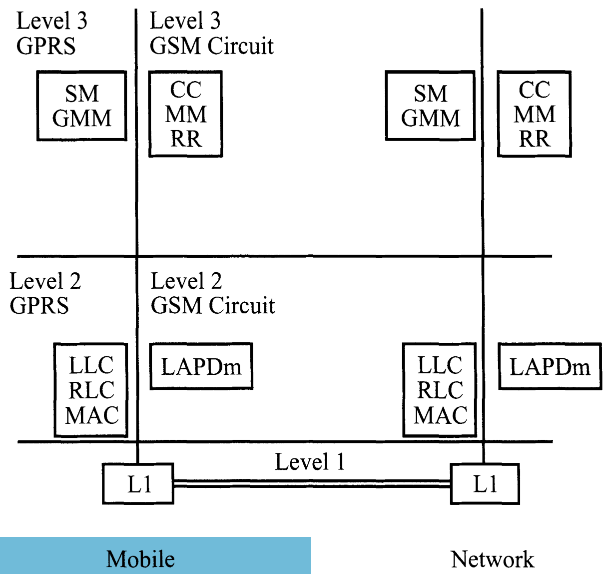


Figure 2.8 General view of the GSM protocols.

data flows originating from one or more users. Level 2 of GPRS is made up of three sub-layers: LLC, RLC and MAC.

The LLC (*Logical Link Control*) sub-layer provides a secured data transport service between the mobile and the SGSN. It also allows for the possibility of encrypting the data transmitted.

The main function of the RLC (*Radio Link Control*) sub-layer is to segment the data packets received from the LLC layer so as to obtain data blocks of a size compatible with the GSM physical radio interface. Conversely, the RLC is also responsible for reassembling the blocks of data received from the physical interface and destined for the LLC layer.

The purpose of the MAC (*Medium Access Control*) sub-layer is to multiplex the different data flows, coming from different calls, on to the physical resources that the GPRS can use, taking account of the relative priorities of these flows. Since the capacity of the shared physical resource may sometimes be less than the sum of the data flows to be multiplexed, the MAC has an arbitration function allowing priority calls to be treated ahead of others in case of traffic congestion.

The level 2 layers for the circuit-switched part and GPRS are thus significantly different. In contrast, the UTMS access network (the UTRAN) offers a unified picture of the radio protocols for circuit-switched and packet-switched applications. As we will see later, this picture is very close to the GPRS level 2.

Level 3 contains protocols for managing mobility and call set-up. Like level 2, level 3 is made up of a number of sub-layers that are different in the circuit-switched and packet-switched cases.

The RR (Radio Resource) protocol is the first of the sub-layers. This protocol is specific to the GSM circuit-switched mode. It is mainly used to manage radio interface resources allocated to the mobile. The RR protocol therefore contains all the procedures associated with the allocation and de-allocation of radio resources, as well as the procedures connected with handover.

Level 3 also contains two protocol sub-layers that are transparent to the access network: *Mobility Management* (MM) and *Connection/Session Management* (CM/SM).

The MM sub-layer has two main functions:

- to handle the mobility of the GSM terminal in idle mode, i.e. to take account of the procedures for changing the location area initiated by the terminal;
- to provide the security-related functions, including subscriber authentication, encryption of the call, and terminal identification (IMEI).

Because the functions of the MM sub-layer as defined for circuit-switched and packet-switched modes are different, two protocols have been defined:

- MM, used in circuit-switched mode;
- GMM (*GPRS Mobility Management*), used in packet-switched mode.

The connection management and session management sub-layers are used for managing the call proper.

The CM sub-layer contains all the functions of call set-up and close-down in circuit-switched mode, plus a certain number of other functions, such as the management of supplementary services (activation, de-activation, interrogation) or the sending and receiving of short messages.

In contrast to the CM sub-layer, the SM sub-layer used with GPRS does not provide a service in connected mode. Rather, it contains the activation and de-activation functions of the PDP (*Packet Data Protocol*), used as a preliminary to every exchange of data in packet-switched mode.

2.5 Call Management

2.5.1 Calls in Circuit-switched Mode

The circuit-switched part of the GSM core network uses a signalling protocol specific to GSM, called the *Mobile Application Part* (MAP), based on the transport layers inherited from the SS7 (Signalling System 7) fixed telephony networks: MTP, SCCP and TCAP (Figure 2.9).

The *Message Transfer Part* (MTP) is a protocol layer that provides higher layers with a reliable service for the transmission of signalling messages.

The *Signalling Connection Control Part* (SCCP) allows for the exchange of signalling information at international level or between two different networks, using a global addressing system.

The *Transaction Capability Application Part* (TCAP) is a protocol for handling transactions between two nodes of the network, independently of the user layer (MAP in the case of GSM).

2.5.2 Special Treatment for Speech

The coding of speech is an important element of cellular telephone services. In fixed telephone networks, the speech is digitised, encoded according to one or other of the two rules used on the public networks (the A and μ laws) and transmitted by 64 Kbps circuits. This data rate results from sampling the speech at 8 kHz and coding the samples on to 8 bits. In order to reduce the data rate on the radio

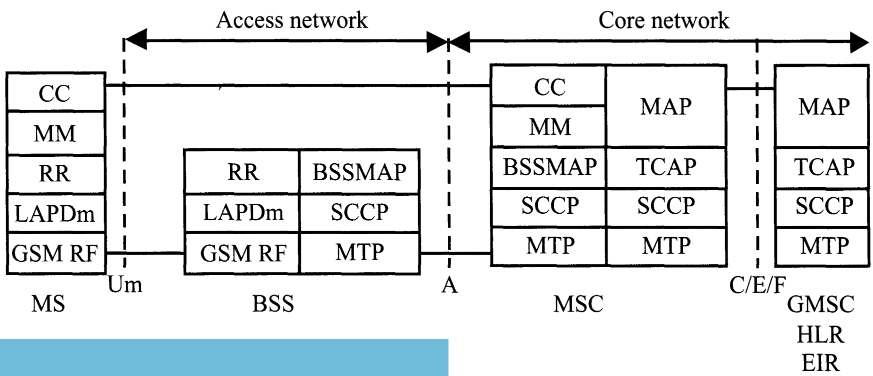


Figure 2.9 The layered architecture of circuit-switched GSM.

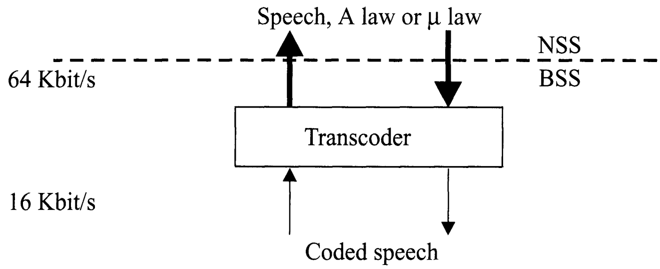


Figure 2.10 Speech transcoding.

interface, a speech-coding function is integrated into GSM terminals and networks. This function allows the data rate for speech to be reduced from 64 Kbps to 13 Kbps (for a coder operating at full speed), while maintaining an acceptable sound quality. Many speech-coding algorithms depend on the principle that successive speech samples are not independent but are correlated; these algorithms reduce the size of the sound packets very effectively (Figure 2.10).

The GSM standard insists that the speech-coding function is situated in the BSS but without stating exactly where, thus giving the network constructor a certain flexibility. It is thus possible to place the coder in the BTS, just in front of the radio interface, or between the BSC and the MSC. The latter is the solution more usually adopted because it allows substantial economies in the transmission to be achieved, speech then being carried on circuits at 16 Kbps in the access network rather than 64 Kbps.

2.5.3 Packet-switched Calls

Unlike calls in circuit-switched mode, the GPRS part of the core network does not use connected (end-to-end) mode transport protocols. GPRS calls in fact use two 'tunnels' (Figure 2.11):

- the first tunnel is used to transfer user data from the terminal to the SGSN;
- the second tunnel serves to transfer the user data from the SGSN to the GGSN, the point of access to the IP network. This tunnel uses the GTP (*GPRS Tunnelling Protocol*), based on UDP (or TCP) transport over IP. A slightly modified version of GTP has been adopted in the UMTS standard.

The use of the term 'tunnel' results from the fact that the data transmitted are encapsulated in a specific protocol (GTP between the SGSN and the GGSN, and SNDCP between the mobile and the SGSN) in order to be transported from one node to another.

From this point of view, the SGSN and GGSN in GPRS play a similar role to the *foreign agent* and *home agent* defined in the Mobile IP framework, i.e. a mechanism developed to provide users with mobility between different Internet or LAN sub-networks.

The GGSN (or home agent) corresponds to the user's reference network; the SGSN (or foreign agent) corresponds to the network visited. When a subscriber

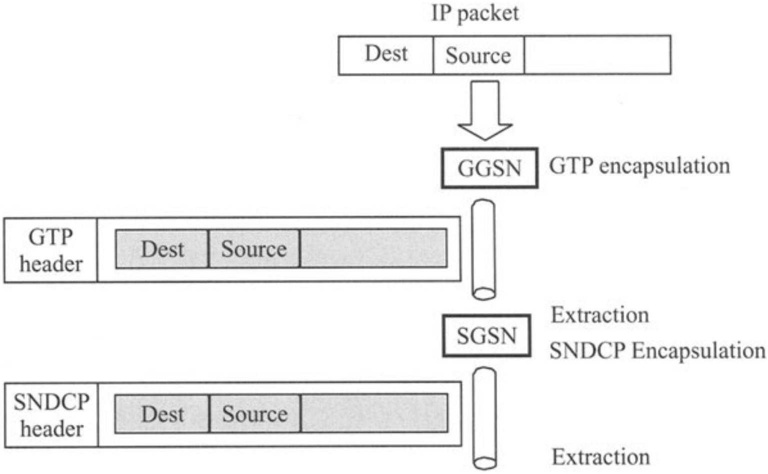


Figure 2.11 Data transmission in GPRS.

to the fixed network sends a packet to a mobile GPRS subscriber, it is routed to the GGSN. Depending on the location of the mobile subscriber, the GGSN will encapsulate this packet in order to route it to the current SGSN of the mobile.

The encapsulation in fact allows the routing of information packets as seen by the two terminals (the routing addresses are the IP addresses of each terminal) to be dissociated from the secondary routing made necessary by the mobility of the user (the secondary routing addresses are the IP addresses of the SGSN and GGSN).

The GTP encapsulation protocol sits on top of a transport mechanism of the TCP/IP or UDP/IP type, levels 1 and 2 not being defined by the standard (they might, for example, be frame relay) (Figure 2.12).

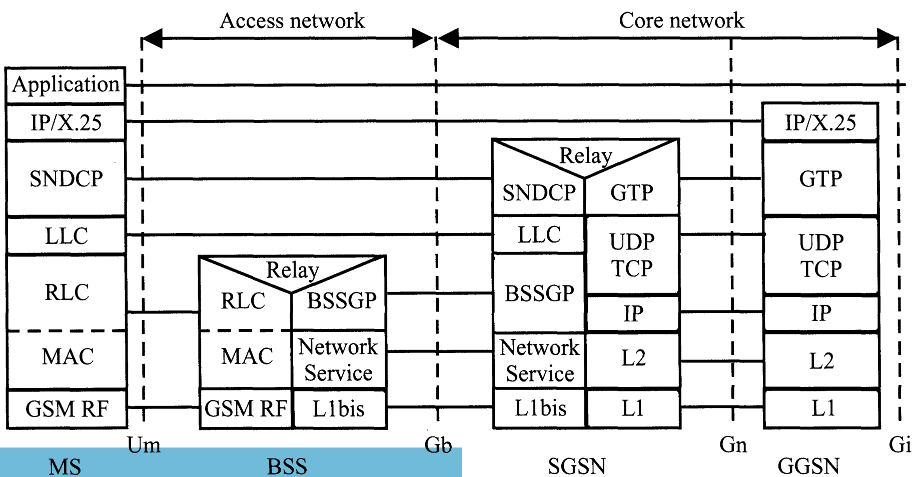


Figure 2.12 GPRS user plane protocols (source: 23.060).

When we come to describe the UTRAN radio interface, we will see that the LLC sub-layer (error detection, re-transmission and acknowledgement of frames, encryption) and the SNDCP sub-layer (which implements the radio encapsulation and also compression of the IP header), although specific to GPRS, have their functional equivalents in UMTS.

2.6 Management of Mobility in a GSM Network

This section presents some basic concepts regarding the handling of terminal mobility within a GSM network. The mechanisms employed are significantly different depending on the nature (circuit-switched or GPRS) of the service used.

As we will see in Chapter 8, the functions and concepts of mobility in the core UMTS network are almost identical to those defined in the GSM standard. As far as the handling of mobility in the access network is concerned, however, new concepts, allowing much more flexibility, have been introduced in UMTS.

2.6.1 Mobile in Idle Mode

When the mobile is in idle mode, i.e. powered up but not engaged in a call, it must choose one, and only one, reference cell. This reference cell is the cell best placed to deliver a service to the subscriber when needed (set up a telephone call, send or receive a text message, etc.).

The characteristics of idle mode in circuit-switched GSM and GPRS are quite similar:

- the mobile is powered up and registered with the network;
- the mobile is inactive, but is in a position to initiate or receive a call, or a packet transfer in the case of GPRS, coming from the network.

In the course of time, if the mobile moves around, it may happen that a change of reference cell becomes necessary. This change is carried out by the mobile autonomously, in a manner determined by parameters furnished by the network and the radio reception criteria broadcast by the cells of the network.

In the case of circuit-switched GSM, cell changes in idle mode are not signalled to the network. Instead, every change of location area must be communicated to the network through a procedure called *location update*. This information is held by the MSC/VLR (Figure 2.13).

A location area is a group, larger or smaller, of adjacent cells. The value of these location areas is that they allow a call to a mobile in the area to be set up quickly.

When the network wants to set up communication with a mobile in idle mode, a *paging message* is broadcast on a special radio channel called the *Paging Channel* (PCH). If the network had no information regarding the position of the mobile with respect to the radio coverage, it would be necessary to send the paging message to all the cells in the network, which would put a very substantial load on the interfaces and components of the GSM network. Thanks to the use of location areas, the network knows the position of the mobile with some precision, which allows it to send the paging message only to a small subset of the cells in the network.

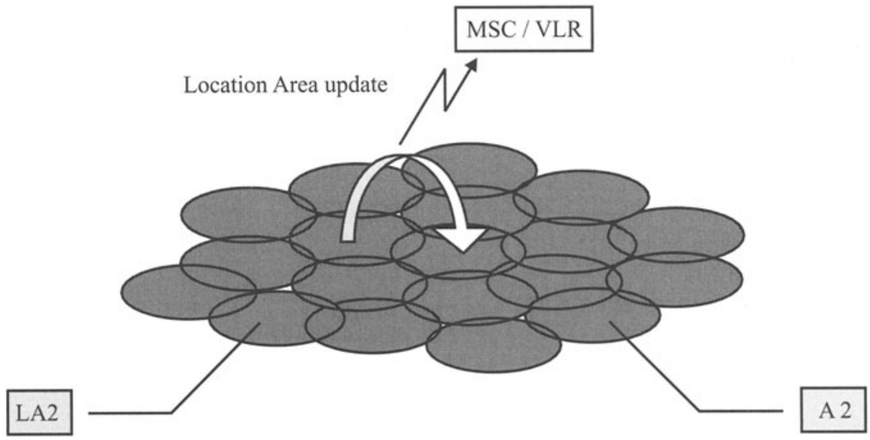


Figure 2.13 Cellular cover.

Unfortunately, this mechanism is not totally satisfactory, for the right compromise on the size of the location areas is difficult to find in zones of high density. Defining small location areas allows the mobile's position to be known with greater precision and thus reduces the loading on the PCH channels. The disadvantage of doing this is that the number of location updates that have to be handled is increased.

From another point of view, it is difficult to work with very large location areas because of the very low channel capacity of the radio PCH.

As we will see in Section 4.4.1, the URA concept (*UTRAN Registration Area*) has been introduced in the UMTS standard in order to alleviate these difficulties encountered by the GSM operators.

So far as the GPRS part of the network is concerned, the principles are very similar. The GPRS equivalent of the location area is the *Routing Area* (RA), the SGSN taking over the role of the MSC/VLR in handling mobility (Figure 2.14). There

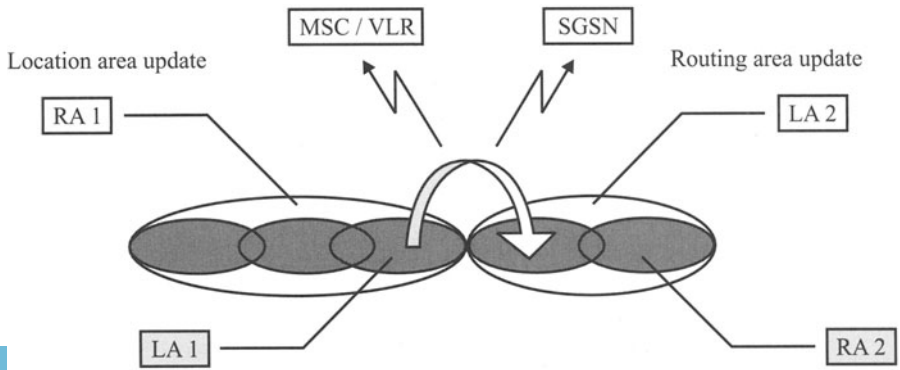


Figure 2.14 Routing and location areas.

is no a priori connection between the location areas and the routing areas, except for the fact that the GSM standard does not allow an LA to straddle two RAs or vice versa.

2.6.2 Mobile in Active Mode

In contrast to idle mode, active mode corresponds to a phase during which the mobile exchanges user data with the network, whether it be transmitting or receiving. Active mode thus corresponds to a telephone call, data transmission in circuit mode, or an exchange of packets in the case of GPRS.

When the mobile is an active mode, a change of cell can still be necessary. In the case of circuit-switched GSM, terminal mobility is completely controlled by the network, depending on parameters chosen by the operator and the effective radio range of the mobile and the network transmission equipment (the BTS) (Figure 2.15).

On the other hand, the GSM standard defines an anchor point in the network that is, by definition, unchanged throughout the duration of the call.

In the case of GPRS, it is possible to handle the mobility of the terminal in either of two modes: either by the network, in a way analogous to the way it is handled in circuit-switched mode, or by the terminal, as in idle mode. The choice of which mechanism to use is left to the judgment of the network (Figure 2.16).

Intuitively, one can easily imagine that, in the case of a very irregular data transfer with low quality of service requirements, the network will prefer mobility to be handled by the terminal itself, in accordance with parameters provided by the network. In this case, the term 'cell reselection' is used.

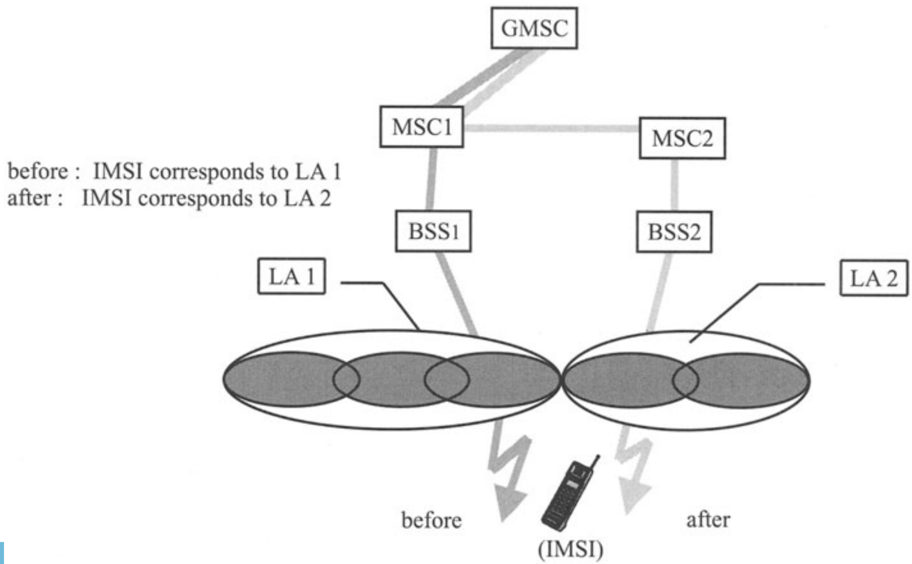


Figure 2.15 Mobility in GSM circuit-switched mode.

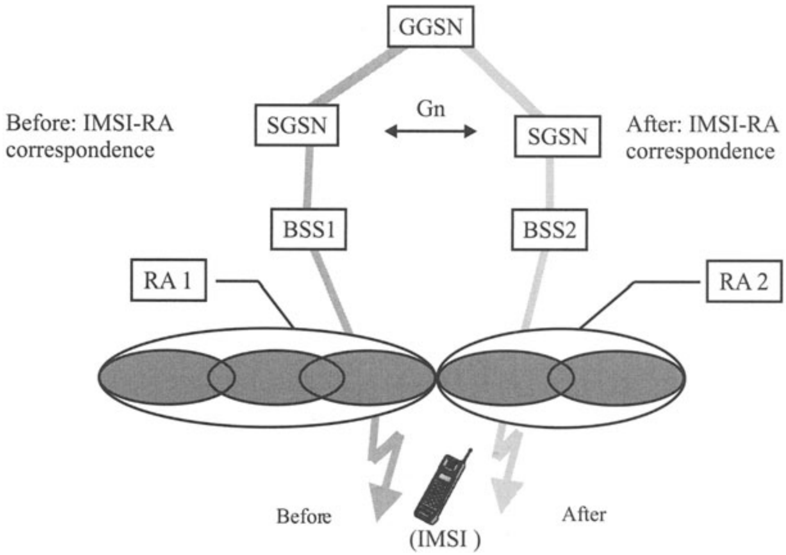


Figure 2.16 Mobility in GPRS mode.

On the other hand, in the case of a transmission with strict quality of service requirements, it will preferable for the network to handle the user's mobility, in order to be able to react more quickly to changes in radio propagation conditions. (The GSM cell reselection procedure is well known for its slowness.)

Cell changing, however, remain inefficient in GPRS, more especially when it is linked to a change of routing area. The main reason for this is the lack of an anchor point in the GPRS architecture. When the mobile passes from one routing area to another, all the data held by the old SGSN relating to the call must be transferred to the new one. Further, the call context in the new SGSN is only created once the mobile has selected its new cell, which further prolongs the interruption in service.

In the case of circuit-switched GSM and, more generally, for services with real-time constraints, the opposite happens. The equipment and target cell resources are allocated and configured before the actual change of cell by the mobile.

The GPRS mobility mechanisms are thus rather inefficient and make it difficult to support packet-switched real-time applications. We will see later that UMTS solves this problem, by offering more efficient mobility mechanisms for packet-switched applications.

3 UMTS Principles

3.1 Basic Concepts

3.1.1 Introduction

As one runs through the ideas and architectural principles of the UMTS standard, one quickly comes to realise that the different 3GPP groups have carried out a significant modelling task. This modelling introduces several concepts that are quite new in comparison with the GSM standard or second generation standards more generally.

There are two main factors that account for this:

- *The services supported*: the application areas for second generation standards, although ambitious for the time, were in fact fairly limited: voice telephony, low speed data circuits, text messaging. Third generation networks have set their sights much higher, following the service developments on the fixed networks. It was therefore necessary to define an initial, more formal framework, capable of supporting the wide range of services to be offered on the wireless networks.
- *The independence of the radio access layer*: in the second generation networks (GSM or IS-95, for example), the radio access layer was defined fairly rigidly and offered little flexibility to the access network. This is particularly true in the case of the GSM or IS-95 circuit-switched data services. The standard for both these services specifies a data format to be used across each of the network's different interfaces, including the radio interface, and imposes its use.

We will see later that UMTS offers a more open architecture and functional decomposition by separating the functions tied to the access technology from those that do not depend on the access mode. This idea of separating the access layer from the rest of the network makes it easier for the UMTS standard to evolve, because it allows the radio access interface to evolve while minimising its impact on the network equipment.

The sections that follow describe the basic concepts of the UMTS standard. These concepts, such as the splitting of the network into *strata* or the idea of the *Radio Access Bearer* (RAB), which will be addressed later, may seem very abstract and of relevance only to a conceptual view of the network. We will see that, in fact, they play a fundamental role in the network structure and the different UMTS protocols and interfaces.

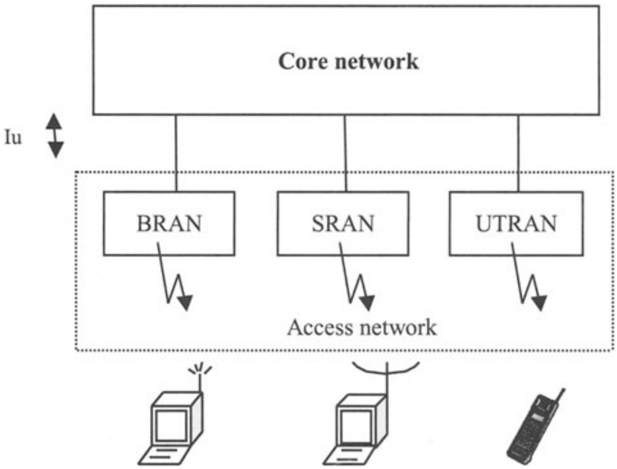


Figure 3.1 General arrangement of a UMTS network.

3.1.2 The Access Network and the Core Network

Like GSM, the UMTS network is made up of a core network and an access network (Figure 3.1). The interface between the core UMTS network and the access network is called the Iu. This interface has been defined in the most general way possible, in order to allow the UMTS core network to be connected to access networks with different technologies, such as:

- the broadband radio access network (BRAN) using an access technology of the WLAN (*Wireless Local Area Network*) type, such as HIPERLAN 2 or IEEE 802.11;
- the SRAN (*Satellite Radio Access Network*);
- the UTRAN, described in this book.

Although the UTRAN is considered the major public technology of UMTS, from a strict standardisation point of view, it is only one of several alternative routes on to the radio highway.

In order to ensure that the Iu interface is independent of the access technology, the concept of Radio Access Bearer (RAB), which allows the communication channel used in the access network to be described generically, is built into this interface. We will return to this concept in Chapter 4.

3.1.3 Decomposition into Strata

In the process of modelling the UMTS network, a decomposition into *strata* has been introduce into the 3GPP specifications. This breakdown, which conforms to the spirit of the OSI seven-layer model, allows levels of independent services in the UMTS network to be separated.

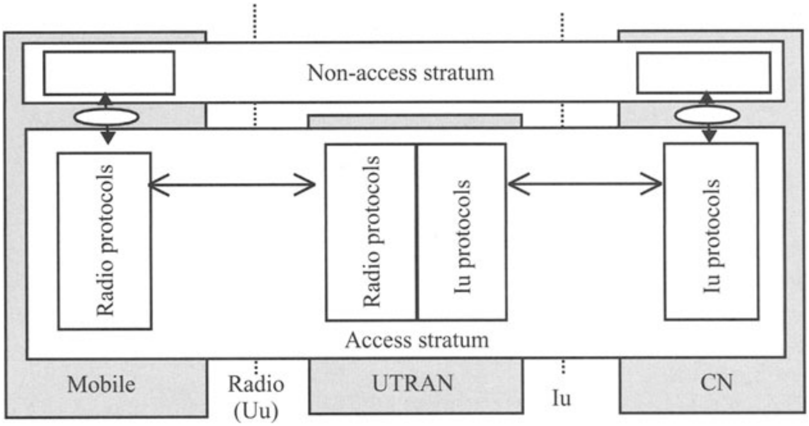


Figure 3.2 Access stratum and non-access stratum.

Access Stratum (AS) and Non-access Stratum (NAS)

In a very general way, a UMTS network is made up of two main levels, called the access stratum and the non-access stratum (Figure 3.2). This decomposition corresponds to a logical breakdown of the functions of the network.

The access stratum brings together all the UMTS network functions connected with the access network, including, for example, handover and the management of the radio resources. By definition the UTRAN, being the UMTS access network, is completely included in the access stratum. But the access stratum also includes part of the mobile equipment (that which handles the radio interface protocols), as well as part of the core network (corresponding to the Iu interface).

The non-access stratum contains all the other functions of the UMTS network, i.e. the ones that are independent of the access network, such as:

- call set-up functions, corresponding to the call control and session management protocol layers for circuit-switched and packet-switched calls respectively;
- mobility management functions for mobiles in idle mode, corresponding to the mobility management and GPRS mobility management layers in GSM.

Table 3.1 shows how the main functions of a UMTS network are split between the two strata.

Table 3.1 Allocation of UMTS functions to AS and NAS.

	Access stratum	Non-access stratum
Management of call signalling		×
Authentication		×
Handover	×	
Management of supplementary services		×
Management of radio resources	×	
Encryption	×	(×)
Compression	×	(×)
Billing mechanisms		×

Some of these functions are included in both strata. This is so in the case of the compression and encryption functions described in the standard as being included in the access stratum but also being able optionally to form part of the non-access stratum.

Links Between the AS and the NAS

The AS acts as a service provider to the NAS. For example, during call set-up, at the request of the NAS, the AS is charged with establishing signalling connections and transmission channels in the access network, according to the type of call and the quality of service attributes agreed between the mobile and the network at the level of the NAS.

A certain number of links between the AS and the NAS, known as *service access points* (SAP), have been defined. These SAPs allow the interactions between the NAS and the AS to be classified according to the nature of the service offered or requested. There are three such access points: *General Control* (GC), *Notification* (Nt) and *Dedicated Control* (DC), see Figure 3.3.

General Control SAP

The *general control* SAP collects together all the services related to the broadcasting of information. It is used, for example, by the NAS to request the AS to broadcast a message right across the access interface. Information broadcast by means of the GC SAP is not addressed to a subset of users but can be read by all the mobiles in the area where the message is transmitted.

Messages broadcast in this way can be of many different types. On the one hand, they may be messages of general interest, such as weather bulletins, information about road traffic conditions, or advertisements. On the other hand, the GC SAP may be used to broadcast network configuration information, such as the present location area.

Notification SAP

Just like the GC SAP, the *notification* SAP collects together services related to the transmission of information on the access interface. The difference lies in the fact that the information transmitted through the Nt SAP is intended only for one or more identified users of the network. The Nt SAP is used, for example, to broadcast paging messages or notification of a conference call. Conference calls,

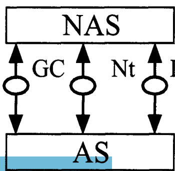


Figure 3.3 Access points between the access stratum and the non-access stratum.

specified in the GSM standard, have not been included in the UMTS standard. This is one of the rare examples of services offered by GSM but absent from the UMTS standard.

Dedicated Control SAP

Unlike the two previous SAPs, the *dedicated control* SAP collects together services intended for a specific mobile. For example, this SAP is used to set up or to close down signalling connections and to exchange information over these connections. Among the list of services supported by the DC SAP we therefore find:

- set-up and close-down of signalling connections;
- the sending and receiving of messages over these same connections. These messages can be of the SMS type (text messages sent or received by the user) or signalling messages coming from the higher signalling layers such as mobility management or call control, for circuit-switched calls, and packet mobility management or session management for packet-switched calls.

The GC, Nt and DC SAPs thus provide a model that allows the services that the AS provides to the NAS to be classified. More concretely, each service request through one of these SAPs will induce exchanges of data or signals on the radio interface or on the access network/core network interface (Iu), via the protocols indicated in Figure 3.4. For example, during an incoming call (i.e. a call coming to a mobile from the UMTS network), the NAS part of the core network will have to use the paging service of the Nt SAP. On activation of this service, the paging message of the RANAP layer is sent by the core network to the access network, then over the radio interface by means of the RRC protocol.

In a similar way, activation of the 'communication channel set-up' service of the DC SAP (also referred to as RAB establishment) will involve exchanges of signalling information at the level of the RRC and RANAP protocols of the AS.

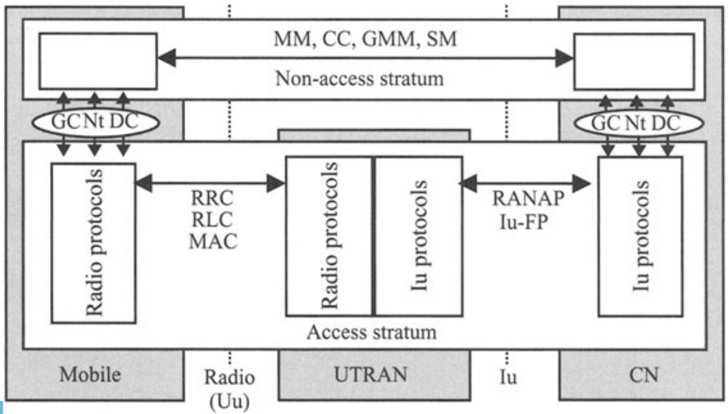


Figure 3.4 Strata, access points and protocols.

3.1.4 RAB Concept

During the setting-up of a call, the type of service and the parameters of the resources that will support it are agreed between the user and the network at the level of the non-access stratum. The access stratum is then charged by the non-access stratum with establishing the communication route within the access network.

The only view that the non-access stratum has of the communication channel used is the *Radio Access Bearer* (RAB). In the access stratum the RAB is decomposed into two parts (Figure 3.5):

- the *radio bearer*, corresponding to the 'radio interface' segment of the RAB;
- the *Iu bearer*, corresponding to the 'Iu interface' segment of the RAB.

By virtue of the principle of the independence of the UMTS levels, the non-access stratum is not aware of the precise characteristics of the RAB, i.e. the way in which it will be implemented in the access stratum. For example, the type of radio channel used, the protocols employed and the way they work are known only to the access network. The RAB is, in fact, characterised only by the attributes, known as quality of service attributes, agreed between the user and the core network.

RAB Attributes

We have seen that, during call set-up, the access network receives a request for the allocation of a RAB, along with a list of the quality of service attributes to be associated with the RAB. The access network must then interpret the RAB attributes and determine the parameters of the radio and Iu resources to be allocated.

In the UMTS standard, the RAB is characterised by the following attributes:

- service class: this indicates the type of service (conversational, streaming, interactive or background) required by the application using the RAB;
- maximum data rate;
- guaranteed data rate;
- SDU size (maximum size or list of sizes) (see Section 4.4.2);

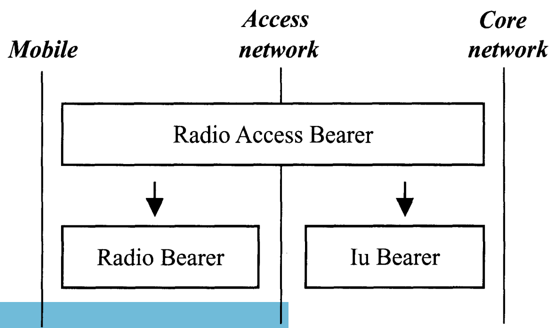


Figure 3.5 Definition of the RAB.

- SDU error rate;
- residual error rate;
- transfer delay;
- priority: this indicates the relative priority of the RAB;
- ability to pre-empt and vulnerability to pre-emption.

Depending on the values of these different attributes, the UTRAN must be able to carry out the following operations:

- Choice of channel coding, i.e. choice of the protection afforded to the user data exchanged over the radio interface. This choice depends on the error rates requested by the RAB.
- Sizing of the radio resources associated with the RAB. Depending on the guaranteed data rate, maximum data rate, service class, and channel coding, the UTRAN determines the performance of the radio resources to be used over the radio interface.
- Allocation of the radio bearer and the Iu bearer. In the case of traffic congestion, the pre-emption attributes and the RAB priority allow the UTRAN to pre-empt an existing resource or to place the resource request in a queue.
- Configuration of radio protocols, depending on the parameters of the SDUs that will be exchanged over the RAB.

Contrary to the GSM position, this process is not described in the standard. The correspondence between the RAB attributes and the configuration of the UTRAN resources is thus left to the implementer, offering network constructors the opportunity of distinguishing themselves from each other by defining their own allocation and configuration algorithms.

3.1.5 Geographical Areas

In an earlier section, we have mentioned the existence of a broadcasting service for data coming from the non-access stratum using the GC and Nt service access points of the access stratum. This service, called the service area broadcast, allows information of general relevance, such as advertisements or traffic information, to be broadcast to a part of the network.

A similar service, the *cell broadcast*, was defined in the GSM standard and implemented by some GSM operators. This service has the disadvantage of making it necessary for the GSM equivalent of the non-access stratum to know about the structure of the cellular coverage. If the cellular coverage is changed, for example because of the introduction of new cells, the GSM operator is forced to reflect these changes in the configuration of certain elements of the core network, which should in principle be independent of problems connected with the access layer.

In order to ensure independence between the cellular coverage of the access network and the areas of interest to the broadcasting services of the non-access stratum of the network, a new concept has been introduced in the UMTS standard, that of geographical zones (Figure 3.6).

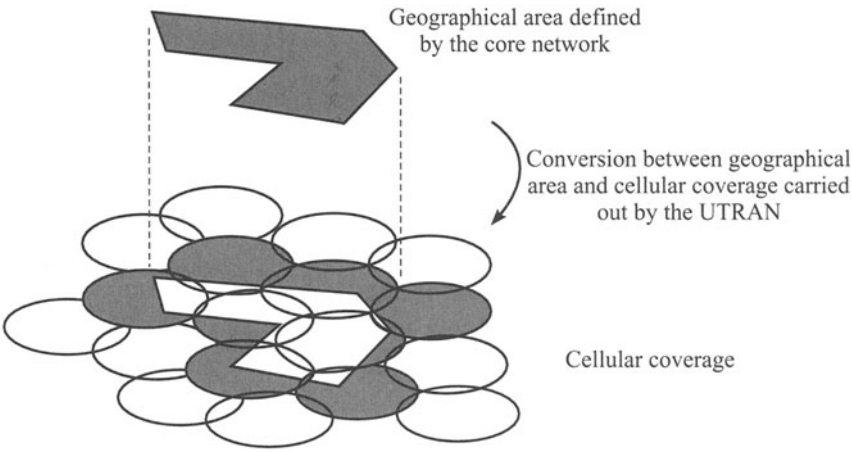


Figure 3.6 Correspondence between geographical area and cellular coverage.

This concept allows, for example, the area to which a message is to be broadcast to be defined not in terms of radio coverage criteria but on a purely geographical basis. The determination of the cells into which the information must be broadcast is a function provided by the UTRAN. Besides simplifying the management of the network, this concept corresponds well with the service needed, because the scope of the information to be broadcast is usually defined geographically.

The location areas and routing areas, which allow the position of mobiles in idle mode to be identified, can also be defined in a geographical manner, contrary to the situation in GSM, where they can only be expressed as cell groupings. The use of location areas is described in Chapter 8.

The standard provides for several different ways of defining a geographical area. In the case of location areas or information broadcasting, the geographical areas are in general defined as polygons whose vertices are identified by coordinates (latitude and longitude) on an ellipsoid of reference. The ellipsoid used as reference by the UMTS standard is that of the WGS (World Geodesic System), which is also used by the GPS (Global Positioning System).

3.2 Architecture and Structure of UMTS

3.2.1 Architecture of the UMTS Network

Figure 3.7 illustrates the general architecture of a UMTS network using the UTRAN access network.

If this schema is compared with its equivalent for GSM, it will be seen that the core GSM and UMTS networks are in reality very similar. The set of components and interfaces of the GSM core network has been retained in the UMTS core network. We will see later, however, that there are functional differences between

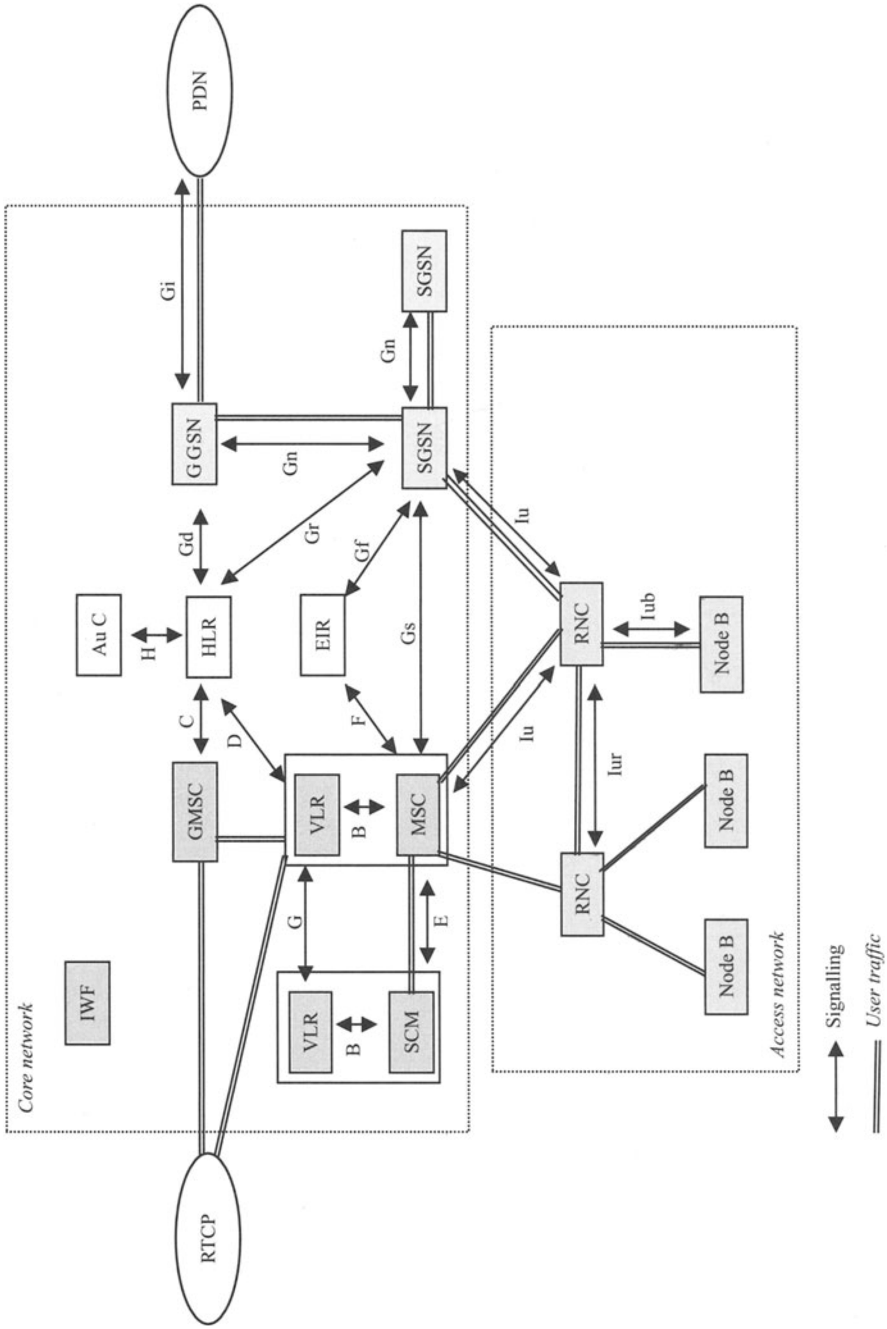


Figure 3.7 Architecture of the UMTS network.

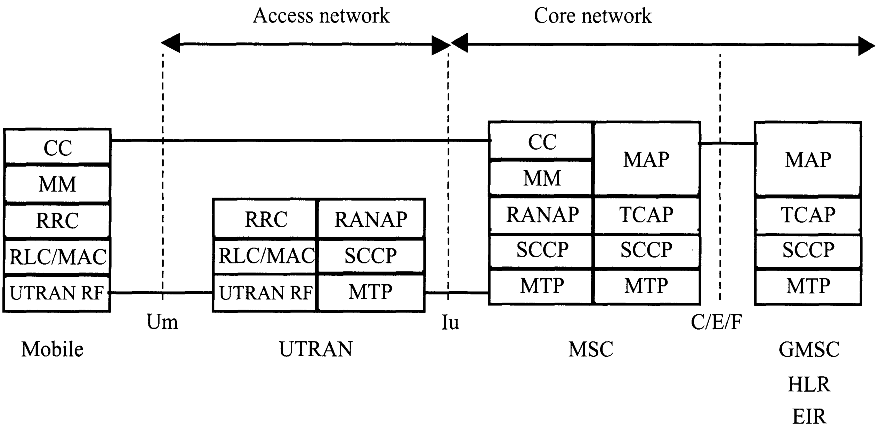


Figure 3.8 Layered architecture for circuit-switched cells.

certain GSM elements and their equivalents in UMTS, in particular with regard to the SGSN and MSC nodes.

The access network, for its part, is made up of the *Radio Network Controller* (RNC), the equivalent of the BSC in GSM networks, and the *Node B*, the equivalent of the BTS. A new interface has been introduced, the Iur, whose function is described in Chapter 8.

3.2.2 Layer Structure of the Network

Figure 3.8 shows the layer structure of the UMTS network for circuit-switched calls. The transport layers TCAP, SCCP and MTP used in the core network on the C, E and F interfaces are identical to those in GSM. The MAP application layer of UMTS is a development of the GSM MAP corresponding to the new services defined in the UMTS framework.

On the access network part, new protocols have been defined. Their role and function are described in Chapter 4.

Figure 3.9 shows the transmission plan for user data in packet-switched services. As for circuit-switched services, the transport layers of the core network between SGSN and GGSN are unchanged.

The signalling belonging to the NAS layer of the network, i.e. CC and MM for circuit-switched calls and SM and GMM for packet-switched calls, is also identical to that used in GSM, with a few extra developments due to the introduction of new services.

3.3 UMTS Core Network

3.3.1 Domain Concept

We have seen that, in the GSM architecture, the core network was split into two distinct parts corresponding to a division between circuit-switching and packet-

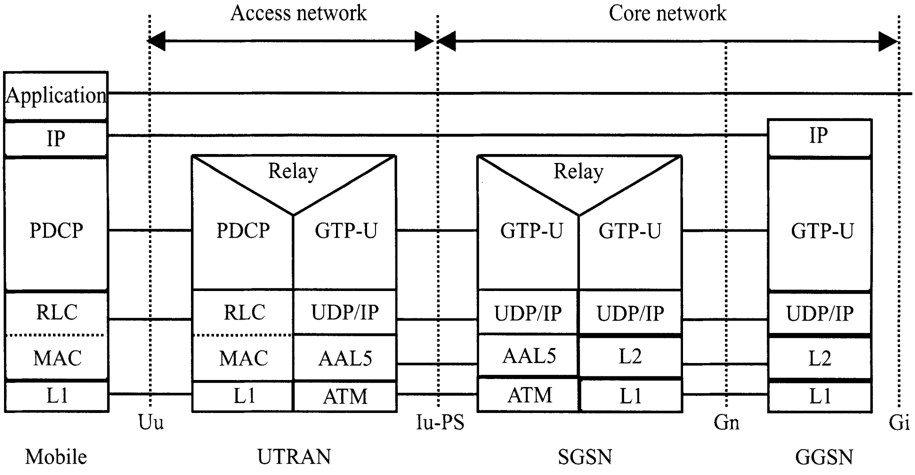


Figure 3.9 The PS user plane (source: 23.060).

switching services. The consequence of this separation is separate management of call set-up and subscriber mobility, which are located in different pieces of network equipment (the MSC/VLR and the SGSN).

This distinction persists in the UMTS standard, with a few changes in terminology. Thus in the 3GPP specifications, the term *service domain* is used. UMTS version 99 defined two domains: the CS (Circuit-Switched) domain and the PS (Packet-Switched) domain (Figure 3.10). The elements of the core network are thus divided into three groups. The first, consisting of the elements of the CS domain, includes the MSC, the GMSC and the VLR. The second group, that of the PS domain, includes the SGSN and the GGSN. The final group includes the network elements that are common to the CS and PS domains: the HLR, the EIR and the AuC.

We saw in Chapter 2 that a class A GPRS mobile is capable of maintaining circuit-switched and packet-switched communications simultaneously. In the same way, a UMTS mobile is able to communicate simultaneously via the CS and PS domains of the core network.

The domain concept is used in a lot of UMTS specifications. It allows the concept of service in the core network to be modelled and makes possible the creation of further service domains in the core network.

3.3.2 Integrated Core Network

Despite the separation of the circuit and packet domains, the GSM standard allowed for the possibility of reconnecting the two domains by means of the optional Gs interface. UMTS goes further in this direction by proposing the concept of an *integrated core network* (i.e. integration of the circuit-switched and packet-switched domains; Figure 3.11). This functional integration simplifies the execution

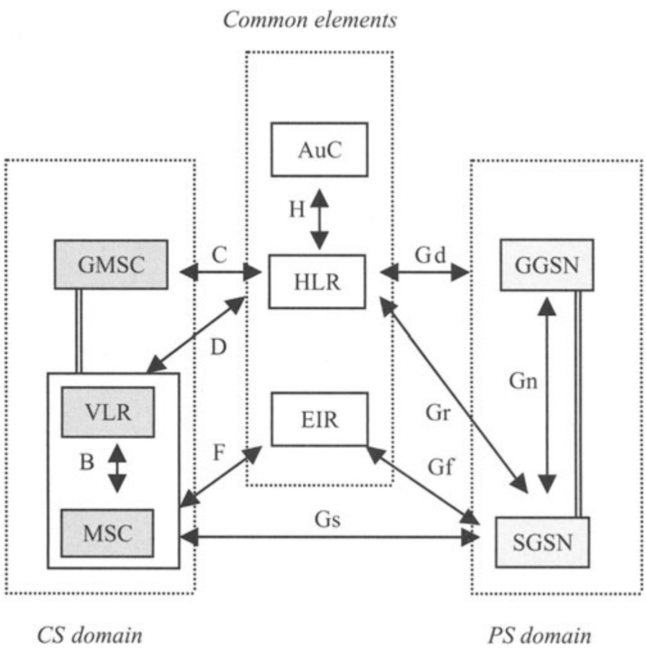


Figure 3.10 Partition of the core network into domains.

of combined procedures for updating the location area in the two domains. The benefits that this brings are twofold: both the time spent in carrying out the updating procedures and the cost of network maintenance are reduced.

The core network entity that brings together the MSC/VLR and SGSN functions is called the UMSC (UMTS MSC).

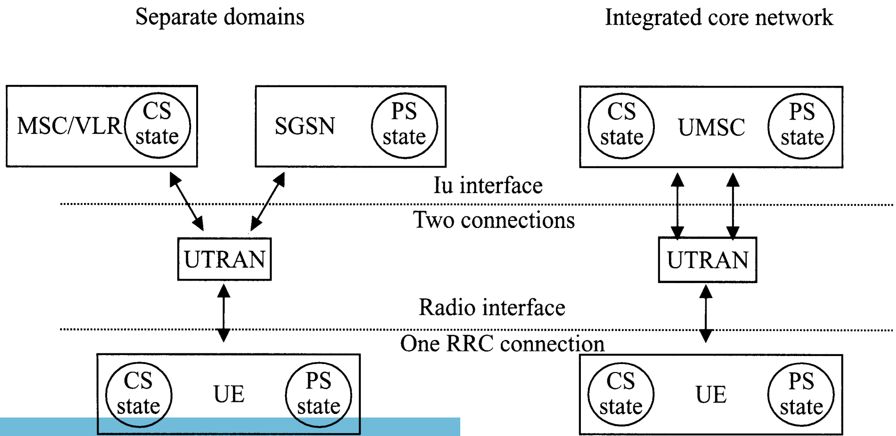


Figure 3.11 Integrated core network.

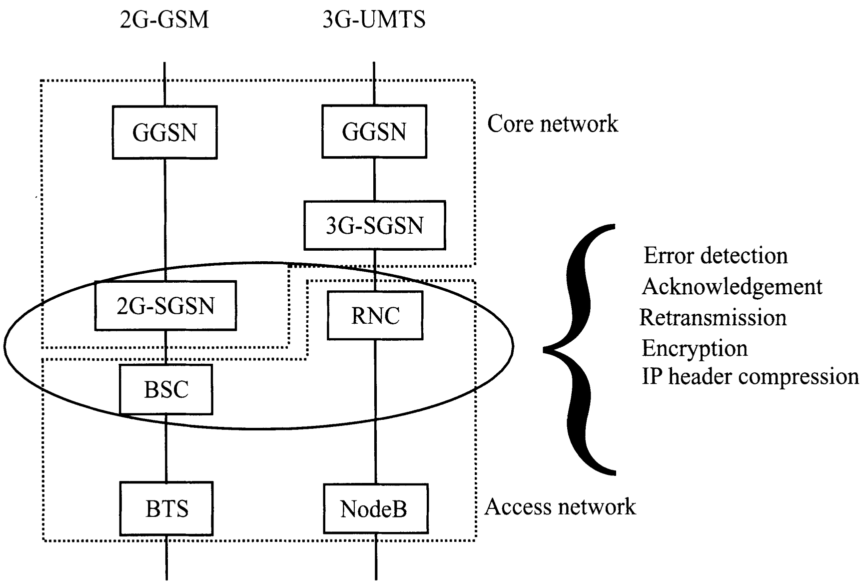


Figure 3.12 Level 2 functions.

3.3.3 Functions Transferred

Level 2 Protocols

In Chapter 2, we saw that the SGSN, situated in the NSS, supported several functions that belong to the radio interface. In order to keep the UMTS core network independent of the access technology, several of these functions, included in the SGSN in second generation networks, have been moved into the access network in UMTS (Figure 3.12). Thus, the functions of the LLC protocol (error detection, encryption, re-transmission and discarding of frames) and the SNDCP (IP header compression) in the GPRS stack have been included in the UTRAN radio protocols supported by the RNC. In Chapter 4 we will see that the functions of the LLC and SNDCP layers of GPRS are provided respectively by the RLC and PDCP layers in the UMTS access network.

Transcoding Function

We saw in Chapter 2 that, in GSM, the transcoder, which carries out the task of coding speech, was included in the access network, its exact position not being specified. (Figure 3.13 shows one of the implementation options, with the transcoder lying between the BSC and the MSC.)

In the UMTS network, the transcoder is situated in the core network, which is the most economical option from the point of view of the transmission network. The Iu interface carries frames of coded speech, which requires a data rate four to ten times less than that on the A interface for voice calls.

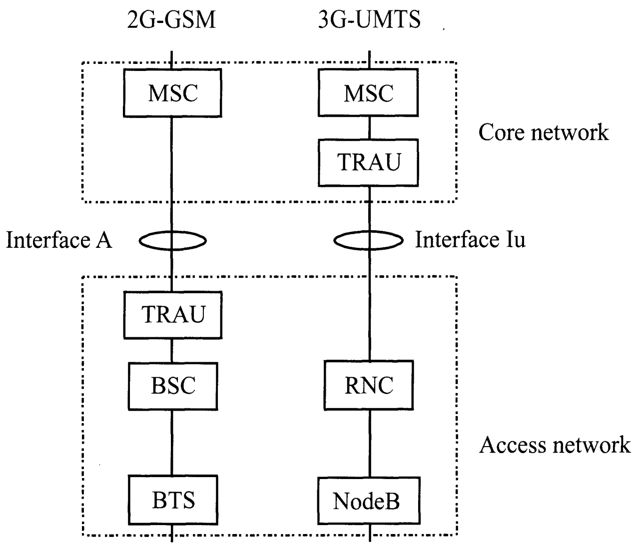


Figure 3.13 The transcoder.

The other result of shifting the transcoder into the core network is to unify the flows of data in the access network. From the point of view of the UTRAN, the telephone service is no longer a special case because the coding of speech is carried out upstream. The transport of frames of speech is, in fact, included in a more general process. In effect, the only thing the UTRAN knows about the different services used by the subscriber is the RAB, characterised by a list of attributes. The RAB for a telephone service is simply one case among many.

We will see in Chapter 4 that the UTRAN comprises a single protocol stack, applied to all supported types of RAB but sufficiently flexible to maintain the quality of service required by each service type.

3.4 UTRAN Access Network

Figure 3.14 shows the overall structure of the UTRAN access network.

3.4.1 Components of the Access Network

Radio Network Controller (RNC)

The UTRAN RNC plays a role equivalent to that of the BSC in GSM networks, i.e. routing communications between the Node B and the core network on the one hand, and the control and supervision of the Node B on the other hand. Due to the existence of a new interface (Iur) in the access network and the definition of new concepts in the UTRAN (in particular, the access and non-access strata), the RNC is functionally quite different from its GSM analogue.

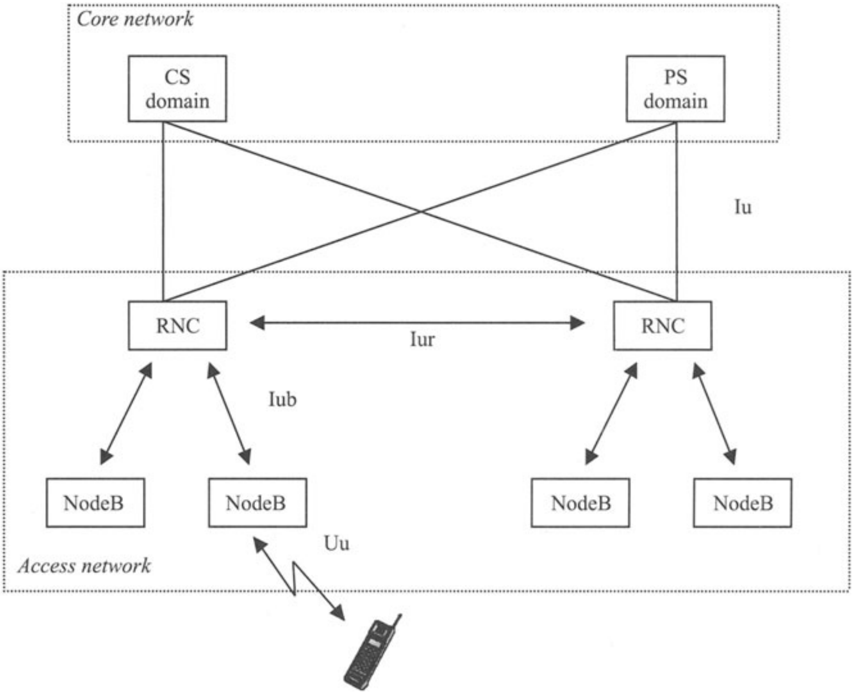


Figure 3.14 General layout of the access network.

Among the new RNC functions are:

- the *relocation* procedure – explained in more detail in Chapter 8;
- management of macrodiversity links – addressed in Chapter 5.

Because of the presence of the Iur interface between the UTRAN RNCs, the specifications identify different types of RNC according to the role they play in each communication (Figure 3.15).

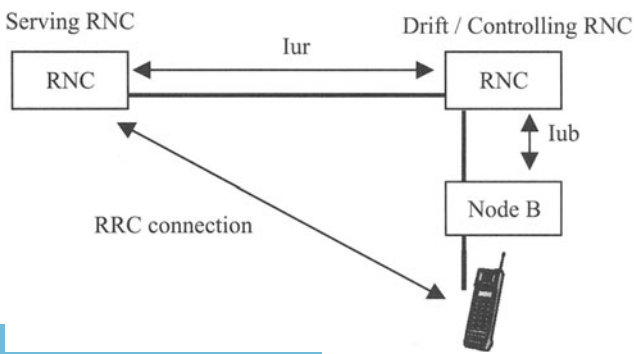


Figure 3.15 The roles of the RNC.

When a mobile is active, a connection RRC is set up between the mobile and an RNC in the UTRAN. We will look into the role of this connection later. The RNC in charge of this connection is called a *serving* RNC (SRNC).

When users move around within the network, they may change cell during a call and may even find themselves in a cell forming part of a Node B that does not belong to their serving RNC. The RNC in charge of these remote cells is known as the *controlling* RNC (CRNC). From the point of view of the RRC, the remote RNC is called the *drift* RNC (DRNC). In the example shown in Figure 3.15, data exchanged between the serving RNC and the mobile travel via the Iur and Iub interfaces. The drift RNC thus plays the role of a simple router as far as this data is concerned.

If each RNC has a well-established role as controlling RNC vis-à-vis the Node Bs that are attached to it, the same is not true of serving and drift roles. Thus each RNC can, at the same time, be a drift RNC for some mobiles and a serving RNC for others.

Node B

The UTRAN Node B is equivalent to the BTS in GSM networks. Its main role is to provide radio reception and transmission for one or more of the UTRAN cells.

The technical implementation and the internal architecture of the Node B are left to the manufacturer. Thus one can conceive of Node Bs made up of one or several cells, using omnidirectional or sectorial antennae (Figure 3.16).

We will see in Chapter 6 that the RANAP protocol (used on the Iub interface and allowing the RNC to 'drive' the Node B) is an open protocol, i.e. it is entirely specified by the 3GPP standard. The aim is to guarantee the possibility of interconnecting an RNC from one manufacturer and a Node B supplied by a different manufacturer.

In order to make it easier to define this protocol, the UTRAN standard specifies a logical model for the Node B. This logical model is independent of the real physical structure of the equipment; it allows the way the RNC 'sees' the Node Bs to which it is connected to be described.

This model is shown in Figure 3.17. It is made up of several components:

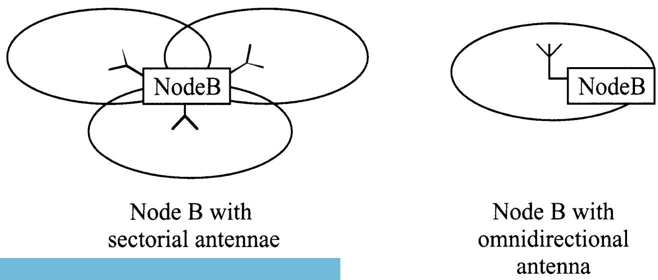


Figure 3.16 Possible implementations of the Node B.

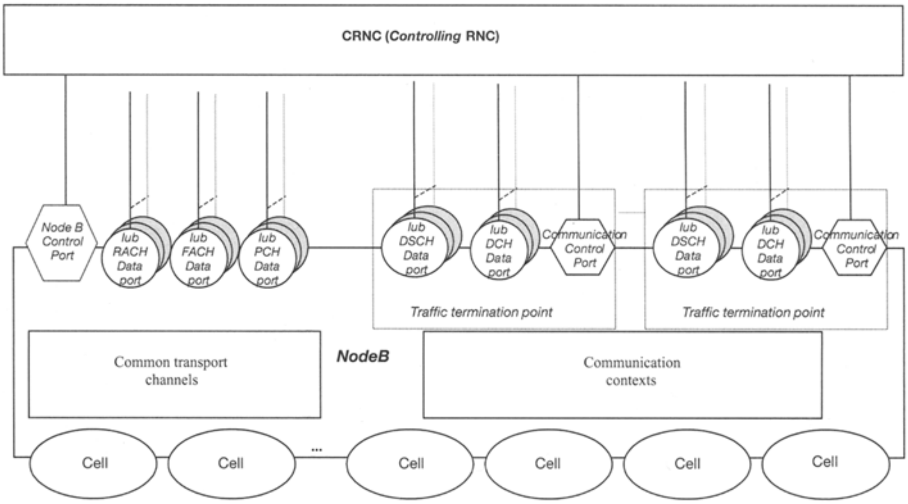


Figure 3.17 The Node B logical model (source: 25.430).

- The *communication contexts* represent the resources allocated to the network users supported by the Node B. Each user has a communication context that includes one or several dedicated channels (DCH) or common transport channels (DSCH: *Downlink Shared Channel*). The management of user resources is carried out through a *communication control port*.
- Each Node B must support a certain number of common transport channels (RACH, FACH, PCH), depending on the number of cells present in the Node B.
- The *Node B control port* is used by the CRNC to configure and initialise the resources supported by the Node B, i.e. the format and structure of the common transport channels, the configuration of the system information that will be sent out on the broadcast control channel of each cell of the Node B, etc.

3.4.2 Access Network Interfaces

The UTRAN access network has the followi ng interfaces:

- the Iu between the RNC and the core network;
- the Iub between the RNC and the Node Bs;
- the Iur between RNCs.

Each of these interfaces supports two types of protocol: application protocols (AP), including signalling exchanges between pieces of equipment, and frame protocols (FP) used to transport user data (Figure 3.18).

In Chapter 6, we will see that the Iu and Iub interfaces are in several ways similar to the A and Abis interfaces of the GSM world. In contrast, the Iur interface has no equivalent in the GSM standard.

The access network interfaces are explained in more detail in Chapter 6.

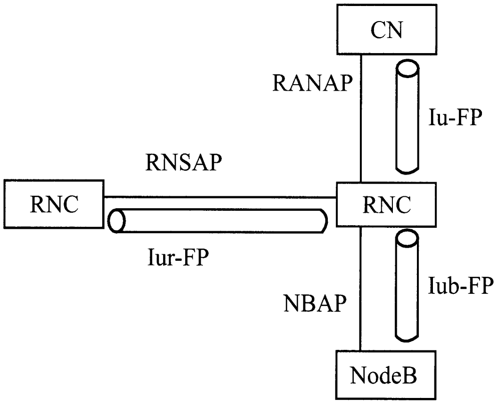


Figure 3.18 General view of the network protocols.

3.5 UMTS Terminal Equipment

3.5.1 Evolution

Within a few years, the cellular telephone market has become a major public market with a substantial level of penetration into the general telecommunications market. When the existing networks make great technological leaps, it is essential that new terminals put on the market are compatible with the different technologies available. Figure 3.19 shows these developments.

The UMTS terminals put on sale in Europe and in other countries where GSM is represented will have to include GSM technology, because this will have much wider coverage than UMTS, at least initially.

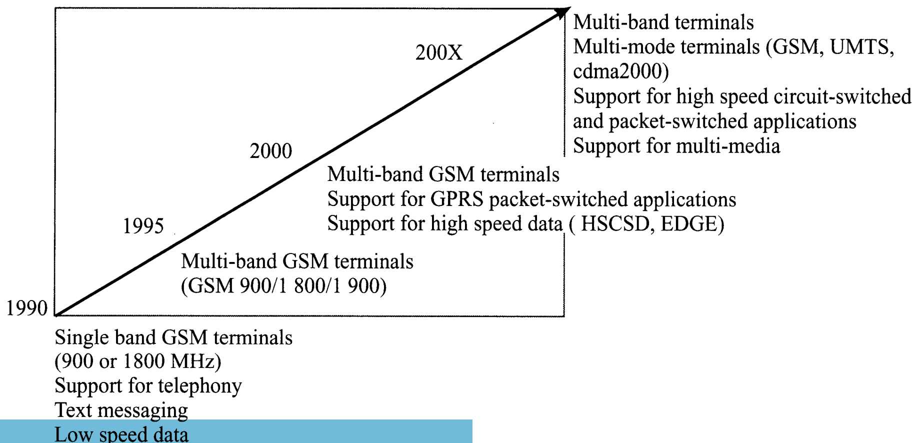


Figure 3.19 The evolution of mobile terminals.

There is one exception to this development model. In Japan, UMTS technology will be deployed without offering continuity with existing systems, Japanese operators having chosen to back a massive deployment in order to get national coverage very quickly.

3.5.2 The USIM Card

As in GSM networks, access to UMTS network services is dependent on the subscriber's smart card, called the USIM (*Universal Subscriber Identity Module*) card, being present in the terminal. In the absence of this card only emergency calls are possible.

The USIM card, like the GSM SIM card, conforms to the ISO/IEC 7816 standard. This standard specifies a number of functional characteristics of smart cards, including for example the dimensions, the position of the contacts, the electrical characteristics, and the data exchange protocol between the card and the terminal.

The ISO 7816 standard defines two card formats: credit card format and plug-in format. The latter is more widely used because of its smaller size (Figure 3.20).

The USIM card contains a number of data items structured into different 'files'. The structure of the data items on the USIM card is in fact an extension of that used on the GSM SIM card.

The USIM card contains all the data relating to the subscriber, including the following:

- the *International Mobile Station Identifier* (IMSI);
- the *Mobile Station International ISDN Number* (MSISDN);
- the preferred language, used for broadcast information and for terminal menu options;
- encryption and integrity keys for the CS and PS domains. These keys are used in the security mechanisms described in Chapter 7;
- the list of forbidden networks;
- the user's temporary identities vis-à-vis the CS and PS domains;
- the identities of the current location area and routing area of the mobile for the CS and PS domains respectively.

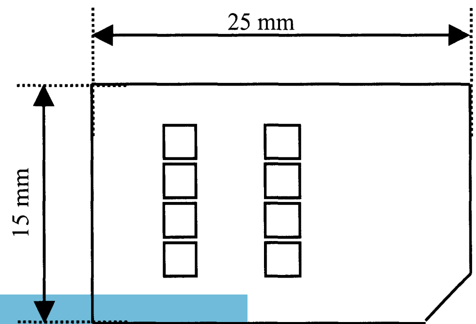


Figure 3.20 The plug-in SIM card.

The standard specifies access and update conditions for each of the data items held by the USIM card. Access conditions, both for reading and for updating, are therefore associated with each elementary file held on the USIM card:

- ALW (*always*) indicates that the data is accessible without restriction;
- PIN (*Personal Identification Number*) indicates that the data is only accessible once the user's PIN has been authenticated;
- ADM (*administrative*) indicates that the data is only accessible to the supplier of the card;
- NEV (*never*) indicates that the data is not accessible.

The access condition ALW will be reserved for the least sensitive information, such as preferred language. In contrast, the IMSI can only be read from the USIM card when the PIN has been authenticated, but can only be modified by the network administrator (ADM).

4 The UTRAN Radio Interface

In this chapter we will present a detailed picture of the UTRAN radio interface and of the communication protocols used on it.

4.1 Multiplexing Techniques on the Radio Network

We saw in Chapter 1 that the UMTS standard offers two different multiplexing techniques on the radio highway: *Time Division Duplex* (TDD) and *Frequency Division Duplex* (FDD) (Figure 4.1).

4.1.1 TDD

In TDD a single frequency is used alternately by the two communication directions (from the mobile towards the network and from the network towards the mobile). This technique is the most flexible when the bandwidth is only available in limited quantities.

4.1.2 FDD

In FDD, each communication direction uses a different frequency. The mobile and the network can thus transmit simultaneously. One of the disadvantages of this technique lies in the 'duplex gap' between the two communication channels, used to separate the radio transmission and reception frequencies. The need to maintain this gap, also known as the guard band, leads to under-utilisation of the radio spectrum.

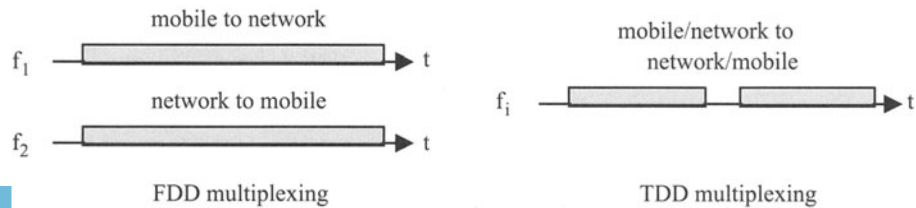


Figure 4.1 FDD and TDD multiplexing.

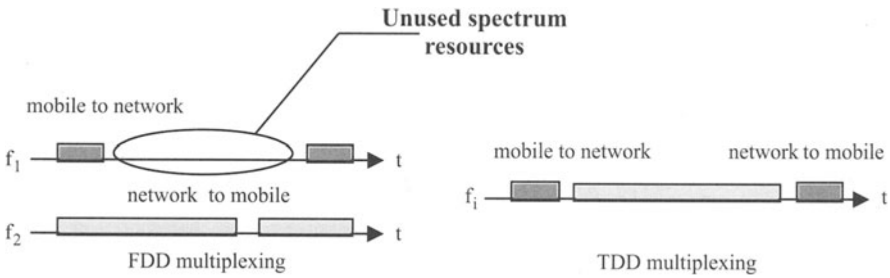


Figure 4.2 Example of asymmetric service.

In FDD, the same amount of bandwidth is generally allocated to the two communication directions, something that is well-adapted to applications with symmetric loading, such as telephony. In contrast, when the loadings are not balanced, which is the case in many data applications such as Internet navigation or message reading, this technique is not optimal.

In this respect, TDD seems better adapted than FDD, for it allows one communication direction to be favoured by using asymmetric resource allocation (Figure 4.2).

In order to avoid a complete separation between the TDD and FDD UTRAN modes in the UMTS standard, it was decided in the 3GPP working groups to make the specifications as independent as possible of the multiplexing technique used. This means that differences between the TDD and FDD modes are almost confined to the physical level of the radio interface.

For several reasons, FDD has benefited from greater support than TDD from different groups within the 3GPP. As a result, the FDD mode is in a more advanced state of maturity in the UMTS standard than the TDD mode. In the remainder of this chapter, therefore, only the FDD mode will be described.

4.2 Radio Interface

4.2.1 Layered Architecture

Figure 4.3 shows the set of blocks that together make up the UTRAN radio interface. These different blocks will be addressed in detail in the sections that follow.

As for GSM, the radio interface protocols correspond to the first three layers of the OSI model.

Level 1 (PHY) represents the physical layer of the radio interface. It implements inter alia the channel coding, interleaving and modulation functions. It is described in detail in Chapter 5.

Level 2 includes the PDCP, RLC, MAC and BMC layers.

The functions that really come under level 2, that is reliable data transmission between two nodes of the network, are provided by the *Radio Link Control* (RLC) layer.

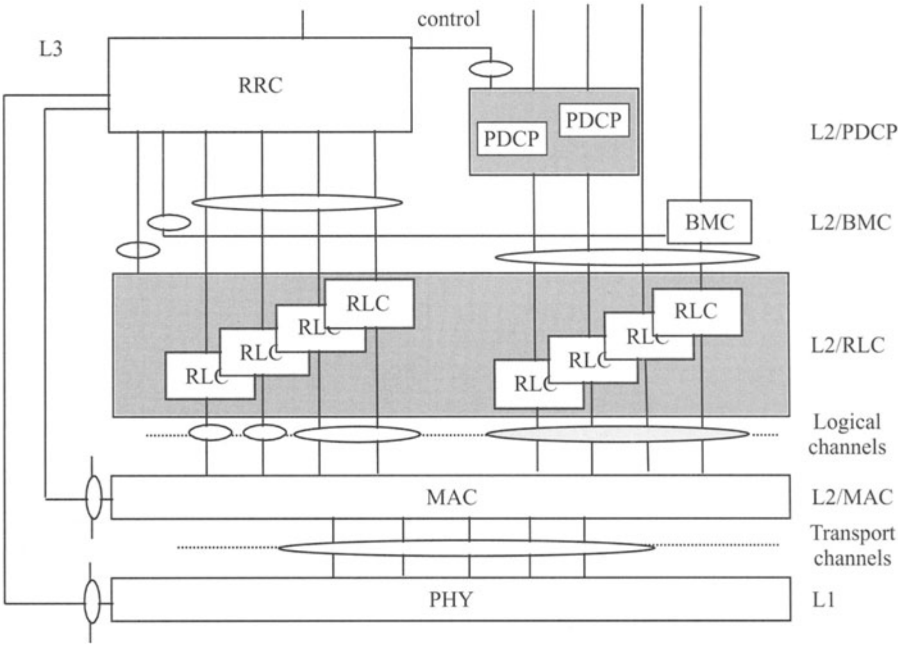


Figure 4.3 Layered view of the UTRAN radio interface (source: 25.301).

The *Medium Access Control* (MAC) layer looks after the task of multiplexing data on the radio transport channel. Specifically, the MAC layer carries out two different types of multiplexing:

- multiplexing different data flows belonging to the same user on a single transport channel;
- multiplexing data flows from different users on a common transport channel or shared resource.

The *Packet Data Convergence Protocol* (PDCP) layer has two main functions. The first is to make the UTRAN radio protocols (in particular, the RLC and MAC layers) independent of the network transport layers; this independence allows network protocols to evolve (e.g. by passing from IPv4 to IPv6) without modifying the UTRAN radio protocols. The second function of PDCP is to compress data and data packet headers, to allow the radio resources to be used more efficiently.

It is also envisaged that, in later versions of the UMTS standard, the PDCP layer will handle multiplexing of several user data flows on to a single logical channel. In the first version of the standard (version 99), the PDCP layer was used only for compressing the headers of TCP/IP packets.

The *Broadcast/Multicast Control* (BMC) layer handles the distribution of messages on the radio interface. The functions of this layer are used to provide the *cell broadcast* services, taken over from GSM into UMTS.

Level 3 of the radio interface contains the *Radio Resource Control* (RRC) layer, the UTRAN equivalent of the RR layer in GSM.

4.2.2 Control Plane and User Plane

The UMTS standard divides the data flows that cross the radio interface into two 'planes': the *control plane* and the *user plane* (Figure 4.4).

The user plane brings together the data that are exchanged at the non-access stratum level of the network. It therefore includes IP datagrams, voice, text messages and information distributed by the non-access stratum of the network.

The UTRAN is transparent at the level of the user plane, in the sense that user plane data is not read or interpreted by the UTRAN. The UTRAN can, in fact, be considered to be nothing more than a transport layer for user plane data.

For its part, the control plane is used to carry all the signalling information between the mobile and the network, through the RRC protocol. There are two types of signalling information:

- Access stratum signalling is exchanged between the UTRAN and the terminal. This signalling relates, for example, to the UTRAN functions of setting an RRC connection or allocation and freeing of radio resources.
- Non-access stratum signalling is concerned essentially with the MM, CM, GMM and SM protocol layers. It provides the call set-up and management functions. This level of signalling is equivalent to the DTAP (*Direct Transfer Application Part*) layer of the GSM standard. It is carried transparently by the RRC protocol.

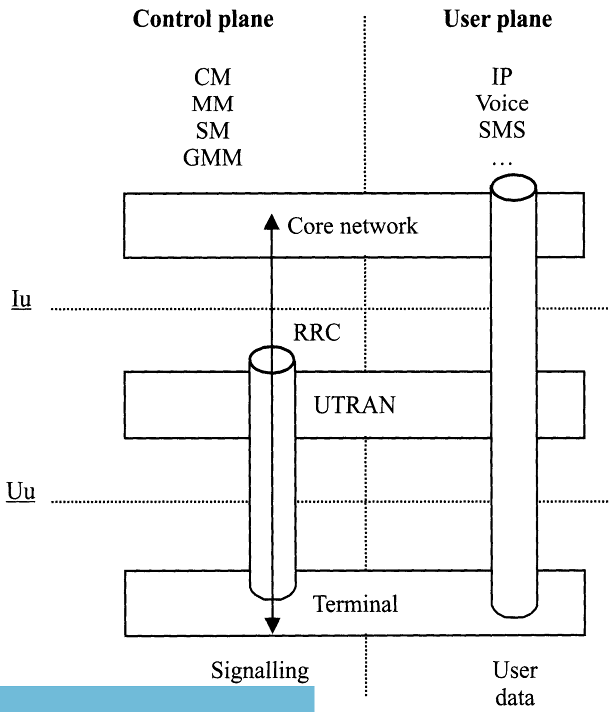


Figure 4.4 The control plane and the user plane.

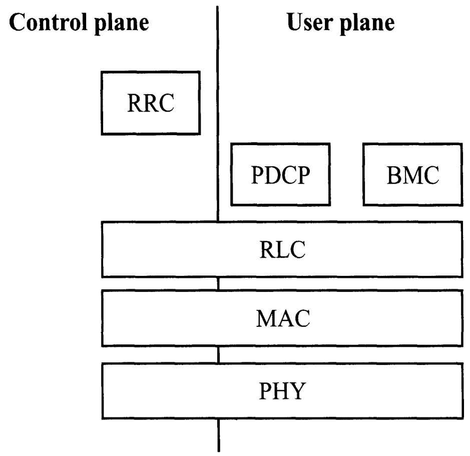


Figure 4.5 Partitioning of the protocol layers according to communication planes.

Figure 4.5 shows how the layers of the UTRAN protocol are split between the control plane and the user plane.

We have already seen that the RRC layer forms an integral part of the control plane. As for the PDCP and BMC layers, they apply only to user plane data. In contrast, the RLC and MAC layers provide services that are applicable to both user plane and control plane data. The two planes are, however, treated separately by the RLC and MAC layers, which allow the radio protocols to work differently for each plane.

4.2.3 Interactions Between the Layers

The model of the radio interface described above shows control links, or interactions, between the layers of radio protocol. These links are used to exchange information between two layers or for one layer to configure another (Figure 4.6).

For example, when a radio interface resource is allocated to a particular user, the RRC layer not only has to send the details of the resource to the user (this is one of the functions of the RRC protocol) but must also provide certain configuration information to the other layers of the radio protocol, such as the mode of operation of the RLC (see Section 4.4.2 below).

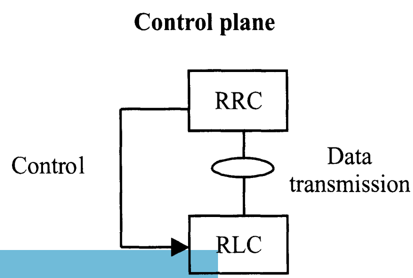


Figure 4.6 Example of the interaction between protocol layers.

Some of these interactions are specified in the UTRAN standard in the form of primitives between layers when the layers are situated in the same node or in the same physical piece of equipment (RNC, Node B). When it is clearly specified that two layers are situated in different nodes (RRC and Level 1, for example), the interactions are specified in the UTRAN network protocol or protocols, i.e. the protocols NBAP and RNSAP on the Iub and Iur interfaces respectively.

4.3 Communication Channels

4.3.1 Introduction

The UTRAN specifications contain a wide range of communication channels, divided into three major classes: logical channels, transport channels and physical channels. These different classes of channel have been created in order to ensure that the different functional levels of the radio interface are independent of each other. The definition of channels belonging to each level makes the UTRAN very flexible, by allowing it to be adapted to the multitude of applications envisaged for third generation networks. The sections that follow describe these different classes of channel.

4.3.2 Logical Channels

Logical channels correspond to the different types of information carried by the UTRAN radio protocols. They are, in fact, the channels offered to the user layers of level 2 of the radio interface.

The concept of a logical channel allows the transmission channel to be considered separately from the use that is made of it. Thus we can conceive of a particular type of transmission channel being suitable for two different uses, i.e. it can support two different logical channels, or it is possible to multiplex two logical channels over the same transmission channel (Figure 4.7).

The number of logical channels in the UTRAN is, in fact, very limited. They correspond to the different types of information flow that travel across the radio interface. The logical channels are divided into two groups: the logical control channels, which are used to transfer information on the control plane, and the logical traffic channels, which serve to transfer data on the user level.

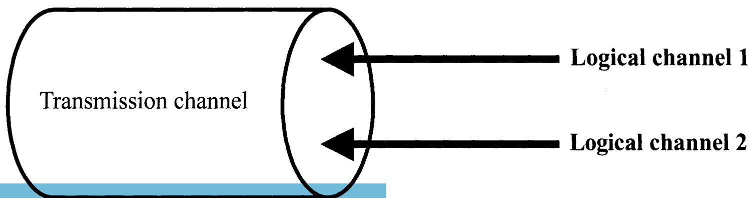


Figure 4.7 Example of multiplexing of logical channels on to a single transmission channel.

Logical Control Channels

The *Broadcast Control Channel* (BCCH) is used for broadcasting control information. The messages broadcast on this channel are known as system information. Amongst other things, they provide the mobile in idle mode with information that allows it to access the network.

The *Paging Control Channel* (PCCH) is used to send paging messages to mobiles on the network.

The *Common Control Channel* (CCCH) is used to send or receive control information for mobiles not connected to the network. In particular, it is used at the very start of call set-up, to exchange the initial signalling messages between the mobile and the network.

The *Dedicated Control Channel* (DCCH) serves to send or receive control information for mobiles that are connected to the network. Almost all the control level signalling therefore travels via this channel, i.e. the UTRAN signalling (the RRC layer) and the signalling of the MM, CC, GMM and SM layers of the core network.

Logical Traffic Channels

The *Dedicated Traffic Channel* (DTCH) is used for the exchange of user data with a mobile connected to the network.

The *Common Traffic Channel* (CTCH) is a unidirectional channel used by the network to send user data to a group of mobiles. The CTCH is particularly used for broadcasting information coming from the non-access stratum of the network, for example, in the case of a service area broadcast.

4.3.3 Transport Channels

Because of transmission problems (interference, fading, masking, etc.), the radio interface is the weak link in the cellular communication network. Depending on the quality of service constraints imposed by the applications supported by the network, it may be necessary to use various mechanisms, such as data-coding systems or other more sophisticated techniques, in order to make the exchange of data over the radio interface reliable.

In the UTRAN specifications, the concept of transport channel corresponds to these different mechanisms. By definition, the UTRAN transport channels represent the data format and, more generally, the way in which information is transmitted over the radio interface. It thus represents the quality of service provided by the network over the radio part of the RAB (i.e. the radio bearer). Each transport channel thus has a list of attributes associated with it, called the *Transport Format Set* (TFS), which defines the data format and the transmission method for transmitting data over the radio interface (Figure 4.8).

The TFS is, in fact, a list of different *transport formats* (TF). This list of transport formats is used by the UTRAN to choose the most suitable format at each moment, in order to use the radio resources as effectively as possible.

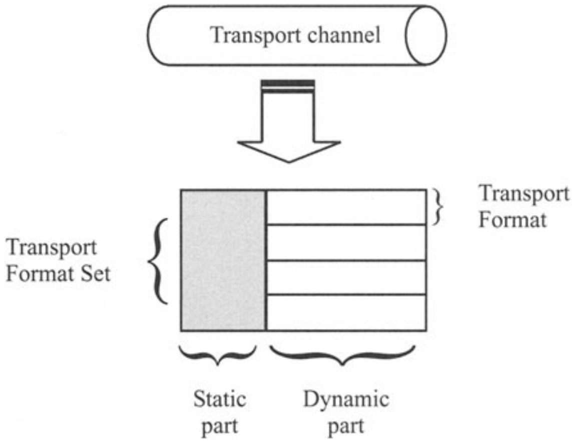


Figure 4.8 Transport channel, TFS and TF.

Each transport format is made up of two parts: a static part, which is common to all the TFs in a TFS, and a dynamic part, which is specific to each TF. The dynamic part of the TF is made up of the following attributes:

- *transport block size*, which defines the size of the blocks of elementary data that will be transported over the radio interface;
- the *transport block set size.*

The semi-static part of the TF is made up of the following attributes:

- *Transmission Time Interval* (TTI), which specifies the periodicity at which a transport block set (or group of transport blocks) is transferred over the radio interface;
- the type of channel coding used (turbocode, convolution code, or unprotected);
- the size of the CRC (cyclic redundancy check), which specifies the number of bits that will be added for detecting transmission errors;
- the channel coding rate (1/3, 1/2), which specifies the number of bits that will be added to protect the data transmitted against transmission errors. A rate of 1/3 means that for each bit of information, three bits will in fact be transmitted over the radio interface.

Figure 4.9 shows an example of a transport channel. The function of the different attributes of the TF will be described in more detail in Chapter 5.

We saw in Chapter 3 that the non-access stratum views the data transmission path in the access stratum as an RAB characterised by quality of service attributes. The choice of the transport format set and the value of the corresponding transport format attributes are determined by the UTRAN on the basis of the quality of service attributes.

The scarcity of bandwidth and the high cost of exploitation licences make the choice of transport channel attributes a critical matter. The difficulty lies in the

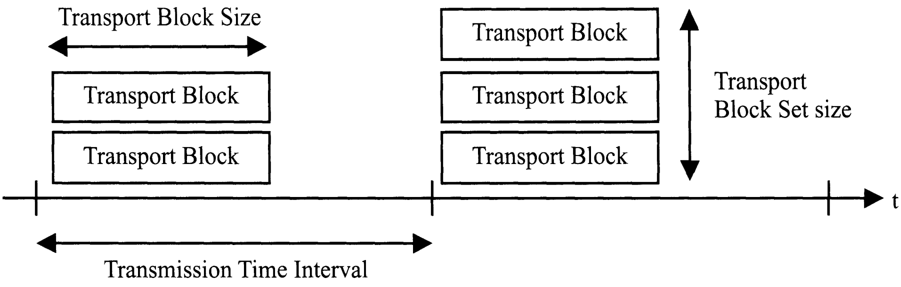


Figure 4.9 Example of a transport channel.

search for a compromise between the quality of service demands of the users and the network capacity, because the decision whether to maximise the quality of service offered to users or to reduce it to the lowest acceptable level has a very great effect on the resource consumption and hence on the network capacity.

Dedicated Transport Channels

The *Dedicated Channel* (DCH) is the only dedicated transport channel. It can be used in an uplink or downlink direction. Because of the separation of the concepts of logical channel and transport channel, the DCH is not typed according to its use. Thus, when the network decides to allocate dedicated resources to a communication between a mobile and the network, the DCCH and DTCH logical channels will each be supported by transport channels of the DCH type, or possibly multiplexed on a single DCH, if their quality of service constraints are compatible.

Common Transport Channels

The *Broadcast Channel* (BCH) is a fixed speed unidirectional (network to mobile) transport channel.

The *Paging Channel* (PCH) is a unidirectional (network to mobile) transport channel.

The *Random Access Channel* (RACH) and the *Forward Access Channel* (FACH) are both unidirectional (mobile to network) transport channels. The *Downlink Shared Channel* (DCSH) is a variant of the FACH, which can be shared.

4.3.4 Physical Channels

We saw in the previous section that the concept of a logical channel allows the type of information flow to be treated independently of the characteristics of the transmission channel. In the same way, the transport channel, specifying how the data is transmitted across the radio interface, is independent of the physical channel actually used. It is thus possible for a physical channel to support different transport channels or for a transport channel to be supported by two separate physical channels (Figure 4.10).

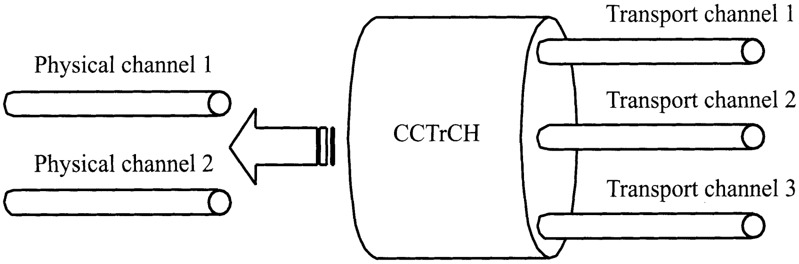


Figure 4.10 Example of the correspondence between physical channels and transport channels.

The *Coded Composite Transport Channel* (CCTrCH) is a concept intermediate between the transport channel and the physical channel. The CCTrCH is the result of multiplexing different transport channels. It can then be supported by one or more physical channels on the radio interface.

The UTRAN standard has defined a number of physical channels, some of which are used only by the physical layer of the radio interface. Only the following channels are able to support transport channels:

- *Primary Common Control Physical Channel* (P-CCPCH);
- *Secondary Common Control Physical Channel* (S-CCPCH);
- *Physical Random Access Channel* (PRACH);
- *Physical Download Shared Channel* (PDSCH);
- *Dedicated Physical Data Channel* (DPDCH).

4.3.5 Correspondence Between Channels

Figure 4.11 shows the correspondence between logical channels, transport channels, and physical channels.

For certain logical channels, BCCH and PCCH for example, the standard offers only a very limited choice of possibilities. The information that travels on these channels is known and is described in the UTRAN specifications. A limited number of transport channels is thus sufficient. On the other hand, in the case of the dedicated channels DCCH and DTCH, the standard offers a large number of possibilities, using either common transport channels (RACH, FACH, DSCH) or a dedicated transport channel (DCH). It is therefore possible to allocate the transport channel best suited to the characteristics of each user's traffic. In the section on radio protocols, we will see that it is not only possible to choose the transport channel or channels when the call is set up or when each RAB is allocated, but it is also possible to change the correspondence between the logical channels in use and the transport channels during the call, depending on the characteristics of the user traffic as it develops.

The correspondence between logical channels and transport channels is handled by the MAC layer of the UTRAN. The standard does not specify how the combinations used depend on the class of traffic; this choice is left to the implementer or

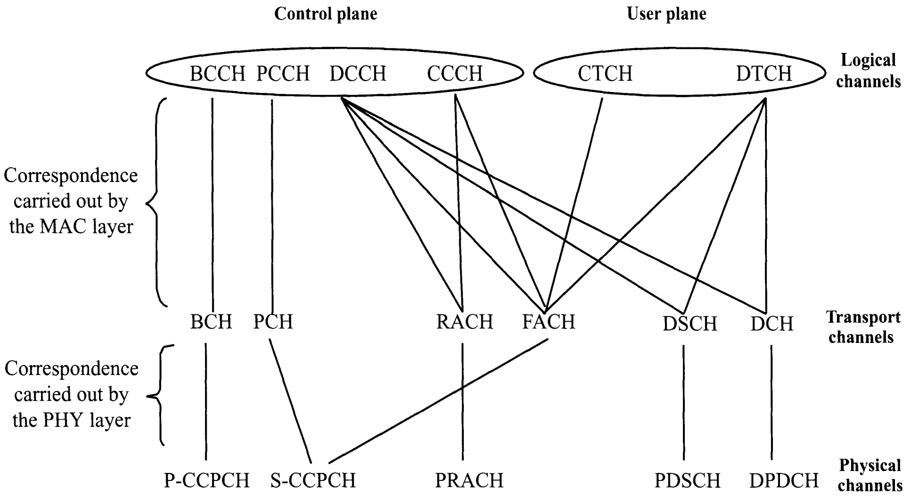


Figure 4.11 Correspondence between channel types.

the network operator. The UTRAN standard specifies only the combinations allowed. This provision allows suppliers to differentiate their product offerings by leaving them the choice of which transport channels to implement, from among the several possibilities offered by the standard.

The correspondence between the transport channels and the physical channels is itself implemented by the physical layer of the UTRAN. This layer enjoys no flexibility in setting up the correspondence, in that each transport channel can only be supported by one type of physical channel.

4.3.6 An Example: Speech Transport

In this section we present a concrete example of the use of transport channels and transport formats in the case of a voice call (Figure 4.12).

We have seen in Chapter 2 that one function of speech coding is to reduce the data rate required from the radio interface. The result of this operation is known as *source code*. Before the frames of source code are transmitted across the radio

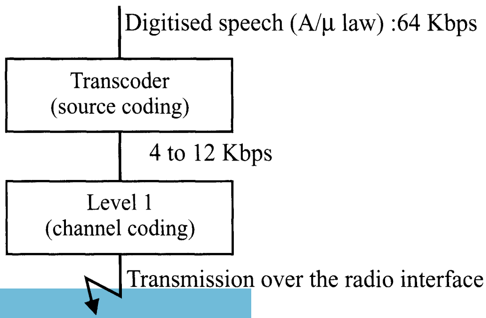


Figure 4.12 Stages in speech encoding.

interface, channel coding is applied, with the aim of protecting the data against transmission errors. Figure 4.12 shows how these different stages are linked. This diagram is applicable both to the mobile and to the network. In the case of the mobile, the two operations are perforce carried out in the same piece of equipment. In the case of the network, however, the source coding operation can be implemented well up-stream of the radio interface (we have seen that, in the case of UMTS, the transcoder was situated in the core network), thus allowing the operator to save transmission capacity on part of the fixed network.

In UMTS, as in GSM, several speech coding methods can be used; they correspond to different frame formats produced by the speech transcoders situated in the terminal equipment and the core network. The network can change the method in use during the course of a call, depending on the radio environment. This is known as *adaptive multi-rate* (AMR) coding.

AMR speech coding is based on the principle that there is no ideal combination of source coding and channel coding. When transmission conditions are good, the number of bits devoted to channel coding can be reduced, to the benefit of the source coding, in such a way as to improve the quality of the speech received. On the other hand, when transmission conditions deteriorate, the number of erroneous frames becomes significant, which has the effect of noticeably lowering the quality of the speech. It then becomes necessary to improve the protection of the data during transmission, which means increasing the number of bits used by the channel coding and so reducing the number available for the source coding. In principle, this entails a reduction in the speech quality but this is largely compensated for by the reduction in the error rate in the speech frames.

Figure 4.13 shows the advantage of AMR coding. When the radio propagation conditions are good (high values of the signal-to-noise ratio, SNR), mode 2 is used. On the other hand, when the conditions deteriorate, i.e. the SNR decreases, using the mode 1 frame format, with a more robust channel coding, allows better speech quality to be obtained.

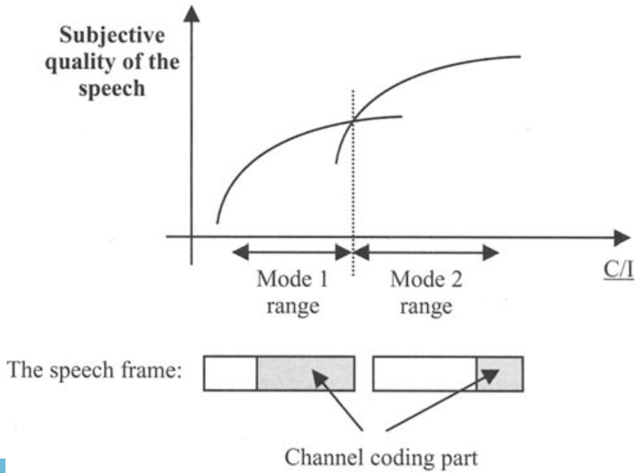


Figure 4.13 Example of two-mode AMR.

Table 4.1 AMR modes.

AMR mode	Source data rate	Class A	Class B	Class C
AMR12-2K	12.2 kbps (GSM EFR)	81	103	60
AMR10-2K	10.2 kbps	65	99	40
AMR7-95K	7.95 kbps	75	84	0
AMR7-40K	7.40 kbps (IS-641)	61	87	0
AMR6-70K	6.70 kbps (PDC EFR)	55	79	0
AMR5-90K	5.90 kbps	55	63	0
AMR5-15K	5.15 kbps	49	54	0
AMR4-75K	4.75 kbps	39	56	0

The speech transcoder divides the bits it produces into three classes, A, B and C, of varying importance. Bits of class A are the most sensitive to transmission errors; they therefore require more robust channel coding than bits of classes B and C.

Table 4.1 shows the AMR modes defined for UMTS. The mode corresponding to the best radio conditions is AMR12-2K, also used in GSM, and known as *Enhanced Full Rate* (EFR). Table 4.2 is an example of a transport channel configuration used for voice transmission. Bit classes A, B and C each require different levels of protection and so a transport channel is dedicated to each class. The special characteristics of each transport channel are contained in the dynamic part of the transport format.

For each speech frame sent on to the radio interface, a specific format (the *transport format combination*, TFC) will be chosen by the network, corresponding to the specific transport format for the transport channel.

Although there are theoretically $8^3 = 512$ transport/format combinations, only eight combinations are really possible, i.e. the ones corresponding to the eight modes of the AMR transcoder. These eight combinations are collectively known as the *transport format combination set.*

The standard defines only the dynamic part of the transport format. In fact, this relates to the size of the frames generated by the transcoder and received by the UTRAN. The static part of the transport format is given here only as an example, the standard imposing no specific transport format on the network builder. In

Table 4.2 Example of a TFS for speech transport.

	Attributes	DCH A (Class A)	DCH B (Class B)	DCH C (Class C)
Dynamic attributes	Transport block size	81	103	60
		65	99	40
		75	84	0
		61	87	0
		55	79	0
		55	63	0
		49	54	0
		39	56	0
	Transport block set size		As above	
Static attributes	Transmission time interval		20 ms	
	Type of channel coding	Convolution		None
	Channel coding rate	1/3	1/2	
	Size of the CRC	8	0	0

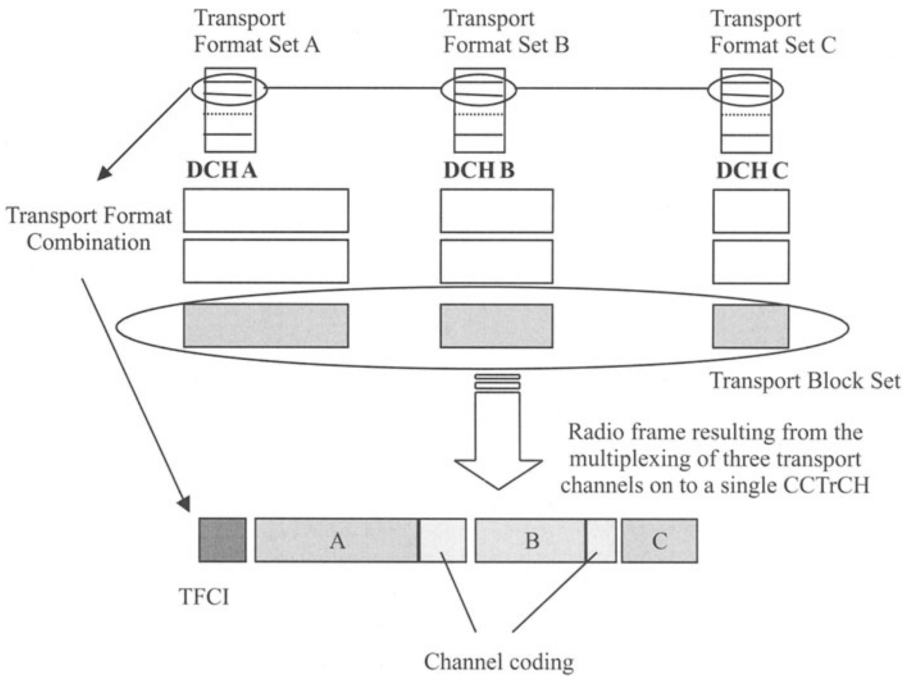


Figure 4.14 General view of transport channel use.

particular, there is nothing to forbid the UTRAN from using only one transport channel instead of three, which amounts to applying the same protection scheme to all three classes of bits.

Figure 4.14 gives a general view of the different transport channels employed in the preceding example. A speech frame is sent to the radio interface every *transmission time interval*. For each of these frames, the UTRAN chooses a *transport format combination* (TFC) specifying a transport format for each of the blocks (*transport blocks*) coming from each transport channel.

After channel coding has been applied to each block, the whole set is then multiplexed on to a single CCTrCH by the physical layer. The TFC chosen by the UTRAN is indicated by the *transport format combination indicator*, thus allowing the receiving entity to decode correctly the data it receives.

4.4 Radio Protocols

4.4.1 RRC Layer

RRC Connection

The main function of the *Radio Resource Control* (RRC) layer is to handle the signalling connection set up between the UTRAN and the mobile. This connection is used during signalling exchanges between the mobile and the UTRAN, for example during call set-up and close-down or to support mobility procedures in the access network.

We have seen in Chapter 3 that UMTS networks offer the possibility of simultaneously exchanging data in both the circuit-switched and packet-switched domains of the core network, i.e. of supporting circuit-switched and packet-switched calls at the same time. There will be, however, only a single RRC connection set up between the mobile and the UTRAN, however many calls are actually in existence between the mobile and the remote terminal or terminals. The RRC layer in effect manages the total set of resources allocated to the mobile, without needing to know whether or not they correspond to different domains of the core network.

The RRC connection is also used to carry signals from the domains of the core network with respect to which the mobile is active (Figure 4.15). In order to separate the different signal flows carried by the RRC connection and to allow the UTRAN to distribute signalling messages coming from the mobile to the correct domain (CS or PS), the RRC layer needs to know the identity of the destination domain.

The UTRAN RRC connection is functionally fairly similar to the RR connection of circuit-switched GSM. However, the fact that the connection is also used for packet services constitutes a break with the way the GPRS works. In the GPRS service, no connection is established between the network and the mobile. Only a communication context, called *PDP context*, is set up in the network; it contains the parameters of the call set up between the mobile and the network. In GPRS, sending data packets on to the radio interface leads to the setting up of a session, called *temporary block flow*, which is deactivated when there are no more packets waiting.

The concept of the RRC connection in the UTRAN allows the handling of circuit-switched and packet-switched calls to be integrated in a single general and unified mechanism. The UMTS network (and, more particularly, the UTRAN RNC) can thus follow the movements of a user by means of the *handover* procedure, in just the same way whether the user is in circuit or packet mode. This point will be addressed in more detail in Chapter 8.

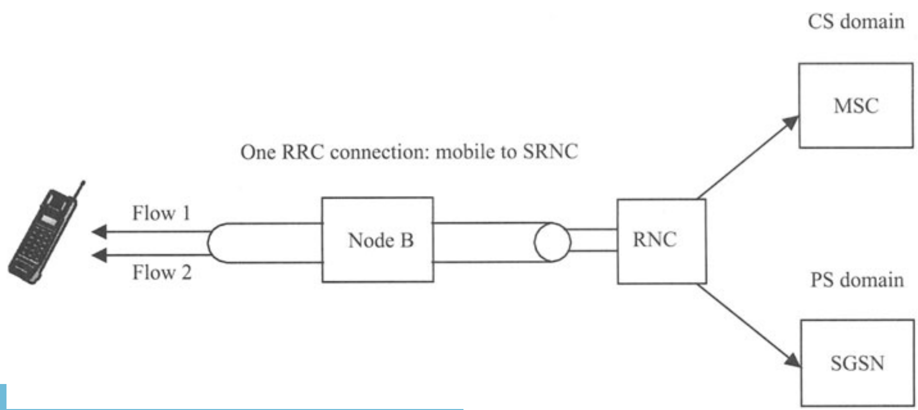


Figure 4.15 An RRC connection and two domain connections.

RRC Connection States

As we have said, in GPRS no connection is established between the network and the mobile; one talks rather of an activated or deactivated session. When a GPRS session enters a phase of quasi-inactivity (which is the case when no user data has been transmitted for a certain period of time), no transmission resource is used in the network, because no dedicated resource has been reserved for the session. The transfer of data from the mobile to the network, or vice versa, can, however, restart at any moment, giving rise to the allocation of temporary resources. This feature of GPRS is relevant to all applications characterised by significant periods of inactivity, such as telemetry or messaging. The allocation of resources on demand thus allows procedures for setting up and closing down network connections, which are long and costly in terms of messages exchanged and CPU time used, to be avoided.

In circuit-switched GSM, the mobile is either in idle mode (i.e. not connected to the network) or in active mode (i.e. connected to the network) (Figure 4.16).

We have seen that in UMTS the RRC connection is present whatever type of service is active. In order to allow the UTRAN to adapt as well as possible to the range of services that must be supported, four different states for the RRC connection have been defined (Figure 4.17). These states constitute a sort of compromise between the two-state RR connection of circuit-switched GSM and the absence of any connection in GPRS.

The states defined by the standard are CELL_DCH, CELL_FACH, CELL_PCH and URA_PCH. These states correspond to different levels of activity of the mobile–network connection. The idle state is the one in which no connection is established between the mobile and the network; it corresponds to the mobile being in idle mode. Transitions between the connected states generally follow from a decision made by the network.

The transition from the idle state to a connected state is only possible at the request of the mobile, regardless of the direction in which the call is being set up.

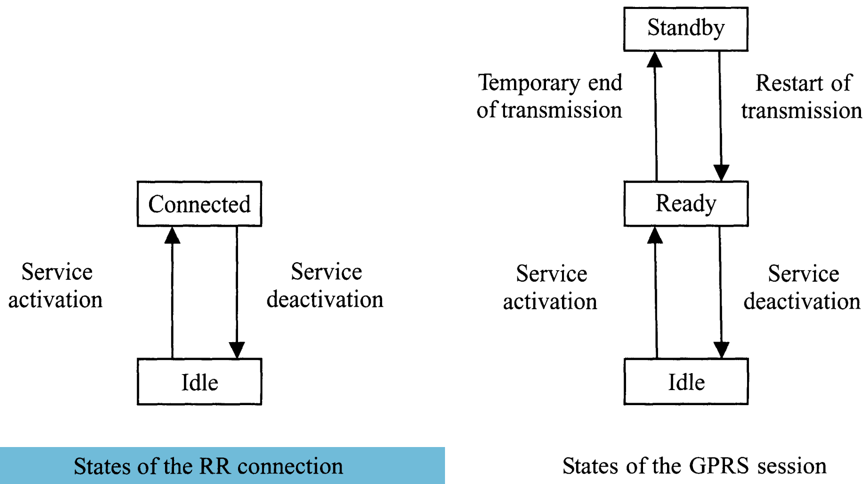


Figure 4.16 States of the GSM mobile.

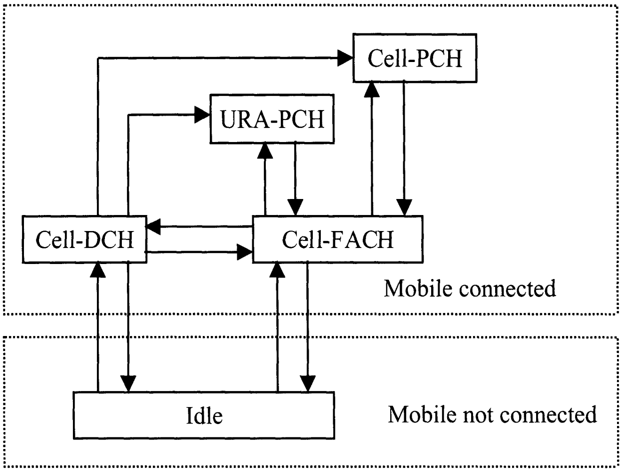


Figure 4.17 States of the RRC connection.

In effect, an RRC connection is requested either by the terminal user wanting to initiate a call or by the terminal responding to a paging message sent by the network (Table 4.3).

CELL_DCH

This state is the one that is closest to the connected RR state in GSM. Dedicated transport channels are allocated to both directions of the call. This is the state that will in general be chosen for applications with real time constraints, such as telephony or broadcast video. In this state, terminal mobility is handled by the network, depending on the actions carried out by the mobile or by the network.

CELL_PCH and URA_PCH

These two states are functionally almost identical. In neither state are any dedicated resources allocated to the mobile and, for this reason, no user data can be transmitted either by the mobile or the network.

When in these states, the behaviour of the mobile is quite close to idle mode. The mobile is, in effect, content to read data transmitted by the network on the BCH and PCH channels. The mobile manages its own mobility in the access network by means of the same mechanisms and criteria as those used in idle mode.

Table 4.3 RRC connection states.

	Uplink transport channel	Downlink transport channel	Mobility checking	Level of activity
CELL_DCH	DCH	DCH	network	+++
CELL_FACH	RACH	FACH	mobile or network	++
CELL_PCH	–	PCH	mobile	+
URA_PCH	–	PCH	mobile	–
Idle	–	–	mobile	NS

The main difference between the two states lies in the knowledge the network has of the position of the mobile within the cellular cover. In the CELL_PCH state, the mobile signals every change of cell to the network. This means that, when the network has to send a paging message to the mobile or transmit user data to it, it can easily reach it over the physical channel PCH of the cell in which it is located. In the URA_PCH state, the network does not know precisely in which cell to find the mobile. Thus, when it wants to reach the mobile, the network has to use other ways of finding it, through the PCH channels of all the cells in the mobile's current URA (UTRAN Registration Area) (Figure 4.18).

The URA_PCH and CELL_PCH bring a new flexibility to the operation of the UTRAN. Depending on what the mobile is doing, on the speed at which it is moving in the network, and on the state of congestion of the radio interface, the UTRAN can decide which of these states to put it into.

CELL_FACH

This is a hybrid state with aspects of both CELL_DCH and CELL_PCH. There is no dedicated channel allocated to the mobile. The mobile (or the network) can, however, transmit user data on the RACH transport channel (or FACH in the case of the network).

The position of the mobile is known almost to the nearest cell. Every change of cell carried out on the initiative of the mobile is signalled to the network, as in the CELL_PCH state. Moreover, in this state the network can ask the mobile to carry out radio procedures with a view to monitoring its movements, as in CELL_DCH mode.

Two Examples

Web Navigation

The data traffic profile of a user navigating the web is made up of two phases:

- a data transfer phase, corresponding to the downloading of a web page on to the user's terminal. From the user's point of view, this is a waiting phase, whose length depends on the size of the resources allocated by the network;

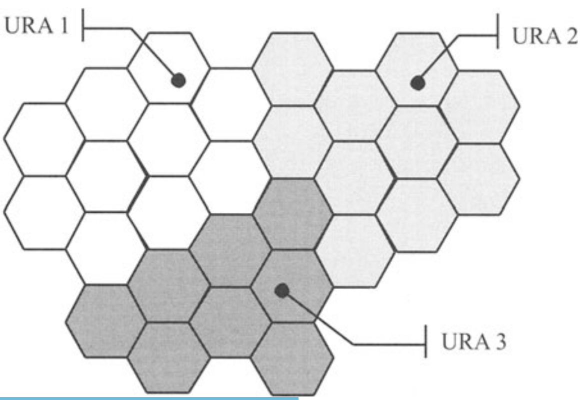


Figure 4.18 The relationship between the URA and cellular coverage.

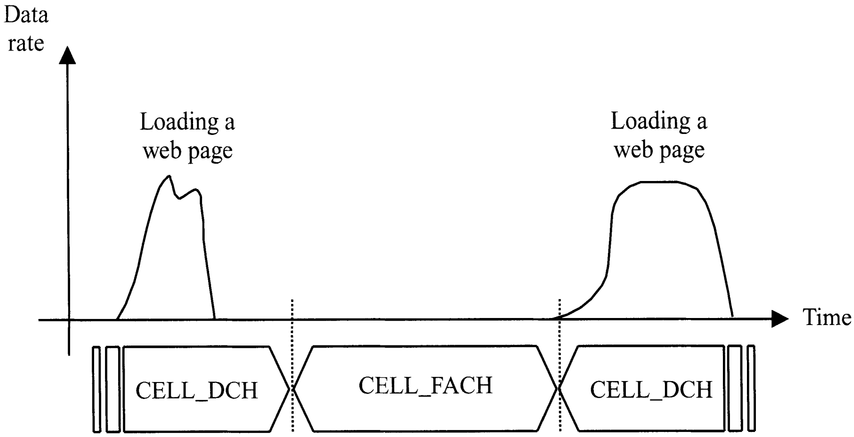


Figure 4.19 Web navigation.

- a phase of inactivity. While the user consults the page, no data is exchanged. During this phase, all resources allocated to the connection are unused.

To guarantee an acceptable quality of service to the user during the data transfer phase and avoid wasting resources during the inactive phase, the RRC layer of the network might decide to use CELL_DCH and CELL_FACH connection states (Figure 4.19).

During the downloading phase, the CELL_DCH state is used, offering a guaranteed level of service so long as the radio propagation conditions are favourable. After a certain period of inactivity, the network would decide to put the connection into the CELL_FACH state; the dedicated resources are freed but the connection is still active.

Tele-surveillance

In this application, a UMTS terminal situated in a house or a flat transmits, every 10 minutes, to a control centre, the state of a number of movement sensors, temperature sensors and fire detectors. The information transmitted amounts to at most a few tens of bytes. To set up and close down a call from the mobile to the network for every transmission could prove costly for the network in terms of the exchange of signalling information.

In this case, it is preferable to set up a connection with the network at the moment that the surveillance system is activated. This connection will remain open throughout the period of the surveillance. The CELL_FACH state will be used for the data transfer and the mobile will then be placed in the CELL_PCH state during the periods of silence (Figure 4.20).

Functions of the RRC Layer

Over and above the management of the RRC connection (i.e. the set-up and close-down of the connection between the mobile and the network and the handling of

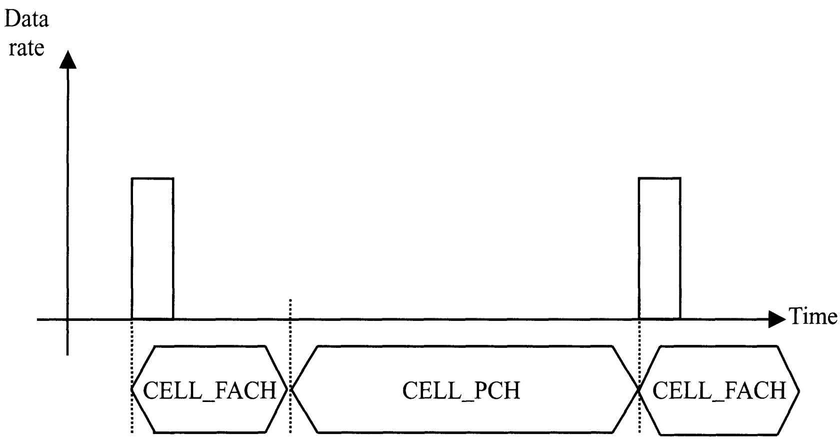


Figure 4.20 Tele-surveillance application.

the different states of the connection), the RRC layer of the UTRAN supports a number of other functions:

- transfer of non-access stratum signalling;
- broadcasting of system information;
- allocation and de-allocation of radio resources. When a connection is being set up, or even during a call, the UTRAN allocates resources to the mobile for both uplink and downlink communication. The parameters of these resources are communicated to the mobile by means of the RRC layer;
- mobility in the access network. The RRC layer includes all the functions needed to keep track of the movement of the terminal in the access network. These functions can be divided into two groups: handling the actions carried out by the mobile and the handover functions.

In contrast to the GSM standard, in which the actions carried out by the mobile and the sequence in which these are sent to the network are defined rigidly, the UMTS standard is extremely flexible. For each mobile connected to it, the network can specify both what it has to do and how often it has to do it.

The handover functions of the RRC layer themselves allow different types of handover to be carried out, both intra-UMTS handovers and inter-system handovers to GSM or cdma2000 networks. These functions will be described in Chapter 8.

4.4.2 RLC Layer

The Radio Link Control (RLC) layer of the UTRAN is the protocol layer that provides the strictly level 2 functions, i.e. reliable transmission of data, coming from the user plane or the control plane, on to the radio interface between the mobile and the UTRAN. We will see in this section that the RLC protocol is very similar to the existing level 2 protocols, such as LAPD or HDLC.

In what follows we will distinguish:

- the *Service Data Units* (RLC-SDU), i.e. the protocol units received by the RLC layer from the higher protocol layers (RRC or PCDP, for example);
- the *Protocol Data Units* (RLC-PDU), i.e. the protocol units provided to the lower layers (in this case, the MAC layer). These units may be prefixed by an RLC header, depending on the RLC mode.

The RLC offers three different modes of working (transparent mode, unacknowledged mode and acknowledged mode), which correspond to the different levels of service offered by the protocol.

Transparent Mode

This mode corresponds to the lowest level of service offered by the RLC layer (Figure 4.21). In this mode, the RLC performs no checking and no detection of missing protocol data units.

The format of the RLC-PDU in transparent mode is extremely simple, since there is no header and it is made up only of a single data field.

Transparent mode can, however, provide one function: segmentation of service data units. Since the PDU has no header in transparent mode, segmentation can only be done according to a pre-established scheme, i.e. by using predefined segment sizes.

Unacknowledged Mode

Unacknowledged mode offers the possibility of segmenting and concatenating the service data units sent (Figure 4.22). The receiving entity is, however, responsible for reassembling the different segments before transferring them to the user layer.

We will see in Chapter 5 that the size of the protocol data units sent to the radio interface is of the order of a few tens of bytes, depending on the capacity of the resource allocated and the parameters of the transport channel used. This order of size is well suited to speech transcoders, which deliver speech frames of comparable size at regular intervals. Thus voice traffic can be carried effectively on the UTRAN radio interface using the RLC transparent mode, since the segmentation and concatenation functions would serve no useful purpose.

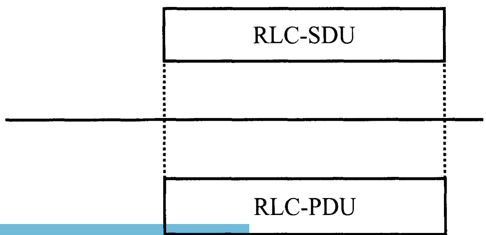


Figure 4.21 Transparent mode.

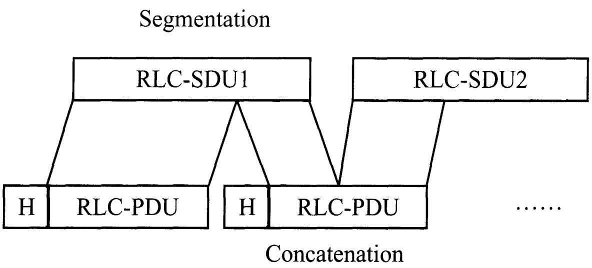


Figure 4.22 Segmentation and concatenation.

In the case of applications supported by the Internet, however, especially file transfer using FTP or web navigation, the IP packets routed by the network are generally of much larger size (from 40 to 1500 bytes). It then becomes necessary to segment the service data units into packets of a size compatible with the demands of the radio interface.

Concatenation allows a partially filled protocol data unit to be filled by all, or part of, the following service data unit. When the last service data unit is sent, the corresponding protocol data unit is, if necessary, filled with padding bytes.

As a complement to the functions of segmentation, concatenation and reassembly, the RLC unacknowledged mode also provides a sequence number check for the protocol data units (Figure 4.23). Thanks to the sequence number contained in the PDU header, the receiving entity can detect missing PDUs, occurring, for example, as a result of transmission problems on the radio interface. When one or more segments of the service data unit received are missing, the SDU is declared incomplete and is not sent to the higher layers.

In contrast to transparent mode, unacknowledged mode requires the presence of a header to provide the functions described above. The PDU header contains essentially the following elements (Figure 4.24):

- sequence number. This field contains the sequence number of the RLC-PDU, used by the receiving entity to detect missing RLC-PDUs;
- one or several field length indicators. These fields allow the receiving entity to separate the RLC-SDUs contained in the data part of the PDU.

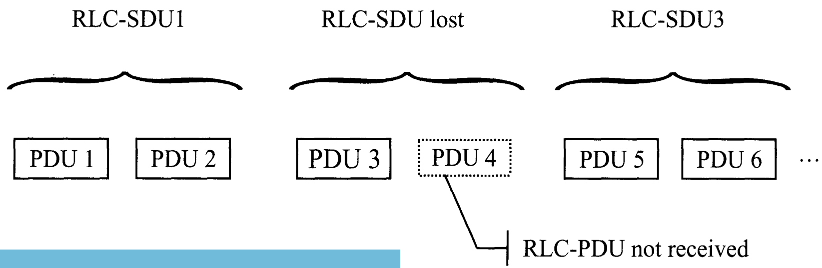


Figure 4.23 Sequence number check.

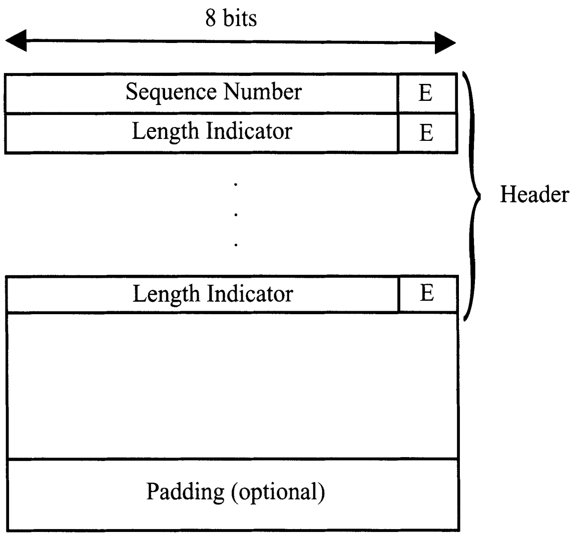


Figure 4.24 RLC-PDU format in unacknowledged mode.

The PDU header is of variable length because it depends on the number of length indicator fields present. For this reason an *extension bit* (E) is used to mark its end. This bit is set to 0 in the last byte of the header and to 1 in all the previous bytes.

Acknowledged Mode

This is the most complex of the RLC transmission modes. In addition to the functions provided by unacknowledged mode, it provides the following functions:

- acknowledgement mechanism for the protocol data units transmitted;
- flow control (through *suspend/resume* messages sent by the RRC layer);
- error correction through re-transmission.

The provision of these functions implies the exchange of control information between RLC entity pairs. This information is, in general, contained in special PDUs called *control* PDUs.

The acknowledgement mechanism employed in the RLC protocol is comparable to that used by the majority of level 2 protocols. The RLC layer uses in effect a sliding window mechanism, according to which a certain number of PDUs, up to the size of the window, can be transmitted without the need for acknowledgement on the part of the receiving entity.

Figure 4.25 is an example of transmission of PDUs from the network to the mobile, which illustrates this mechanism. At the start of the transmission, the window is empty. It is filled as the transmission of the PDUs proceeds. When the window is full, the sender demands, by setting a polling bit (P), that the receiver acknowledges the PDUs sent. The receiver responds by means of a control PDU of type STATUS, containing the sequence number (SN) of the last PDU received

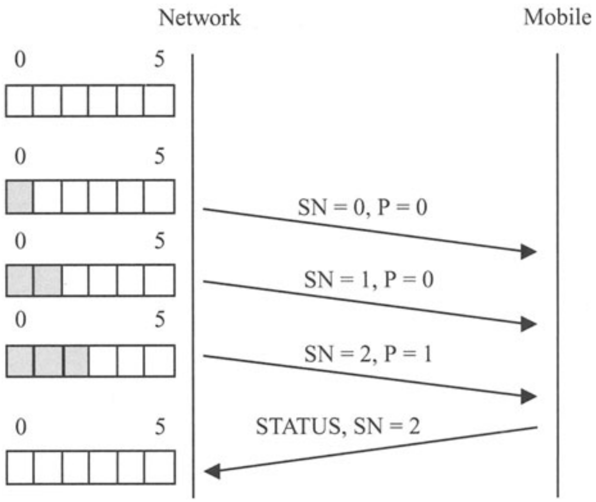


Figure 4.25 The acknowledgement mechanism.

correctly. The STATUS PDU implicitly tells the sender that all the PDUs of lower sequence number have been correctly received.

The size of the window can vary from 0 (every PDU must be acknowledged by the receiver) to $2^{12} - 1$ (the maximum sequence number for the RCL-PDU in acknowledged mode). This value is negotiated between transmitter and the receiver when the link is set up. It can also be modified by the receiver in the course of the transmission.

The receiver can request the re-transmission of certain PDUs that have not been correctly received, by the use of a mechanism comparable to the *selective reject* of the HDLC protocol. As in unacknowledged mode, the missing PDUs are detected by the receiver as a result of a discontinuity in the sequence numbers. The RLC receiving entity can then ask the sender to re-transmit the missing PDUs by means of a STATUS control PDU. This mechanism is illustrated in Figure 4.26.

Figure 4.27 shows the format of the RLC-PDU in acknowledged mode. The sequence number field (coded here on 12 bits), the length indicator and the extension bit (E) of the header are identical to those of the PDU in unacknowledged mode. The additional fields of the RLC-PDU in acknowledged mode are the polling bit (P), used to force acknowledgement of the PDUs by the remote entity, and the *data control* (D/C) bit, which serves to distinguish control PDUs from PDUs containing user data.

For reasons of efficiency, acknowledged mode has provision for including a control PDU of STATUS type as a replacement for the optional padding field at the end of a PDU.

Transmission Modes and Logical Channels

Table 4.4 gives a list of the RLC transmission modes possible with each of the UTRAN logical channels.

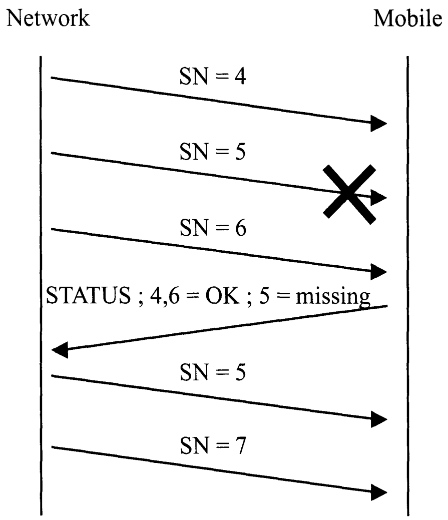


Figure 4.26 Re-transmission of RLC protocol data units.

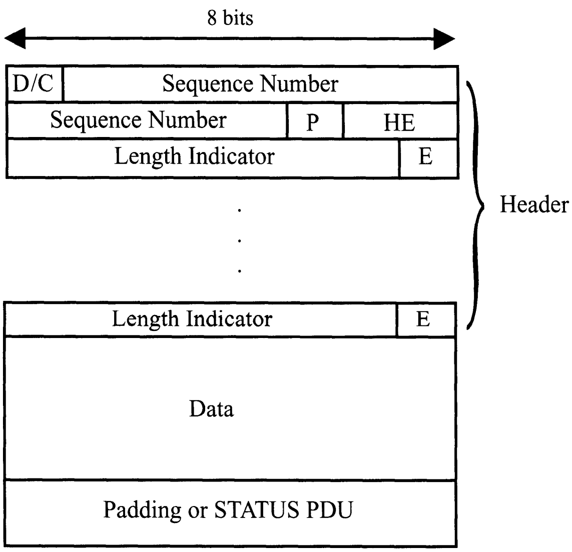


Figure 4.27 RLC-PDU format in acknowledged mode.

Table 4.4 Permitted modes.

	BCCH	PCCH	CTCH	CCCH	DCCH	DTCH
Transparent mode	×	×		×	×	×
Unacknowledged mode			×	×	×	×
Acknowledged mode					×	×

The logical channels BCCH and PCCH use transparent mode, because they are unidirectional channels used by the RRC layer. When the size of the blocks to be transmitted is greater than that of the transport blocks, segmentation and reassembly are carried out by the RRC layer. The dedicated logical channels DCCH and DTCH are the only ones that can use all three RLC modes.

4.4.3 MAC Layer

The main function of the medium access control (MAC) layer is to control access to the radio network. This function is implemented by means of the following two sub-functions:

- multiplexing data on the transport channels;
- choice of transport channel and of the transport format, i.e. the format of the data transported.

Multiplexing

Because of the great flexibility of the UTRAN, the MAC implements different types of multiplexing.

When a common transport channel is used (the FACH, for example), the MAC layer multiplexes the data flows of the different users of the channel. Further, the MAC layer can multiplex different logical channels belonging to the same user (a DCCH and one or more DTCH) on to a single dedicated transport channel of DCH type. Figure 4.28 illustrates this double multiplexing function.

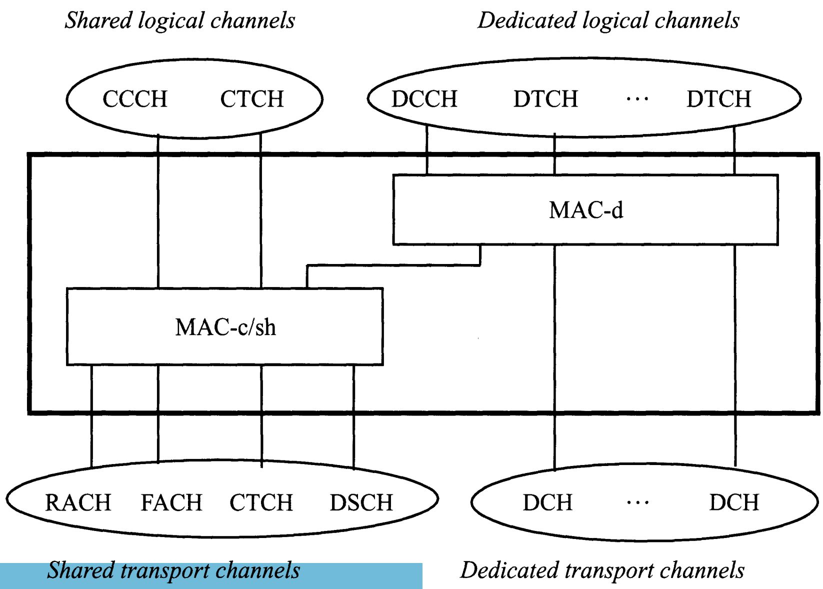


Figure 4.28 The MAC layer.

Multiplexing the dedicated logical channels belonging to a single user is carried out by the MAC-d (*MAC dedicated*) function. The data flows are either directed towards one or more dedicated transport channels (DCH), or are multiplexed on common transport channels, along with data flows coming from (or destined for) other users. This second multiplexing function is implemented by the MAC-c/sh (*MAC-common/shared*) part of the MAC layer.

Choice of Transport Format

In the section on transport channels, we have seen that a list of formats that can be used, the TFS (*Transport Format Set*), is associated with each transport channel. The choice of the data transport format is made by the MAC layer for each transmission period. Based on the capacity offered by the resource allocated, the MAC chooses the optimal format for each transport channel.

The MAC Protocol Header

Depending on the functions provided by the MAC layer, such as the type of multiplexing carried out or the type of transport channel used, a protocol header may be necessary. Figure 4.29 shows the format of the MAC layer PDU. The data field in the MAC-PDU contains the whole of the PDU received from the RLC layer.

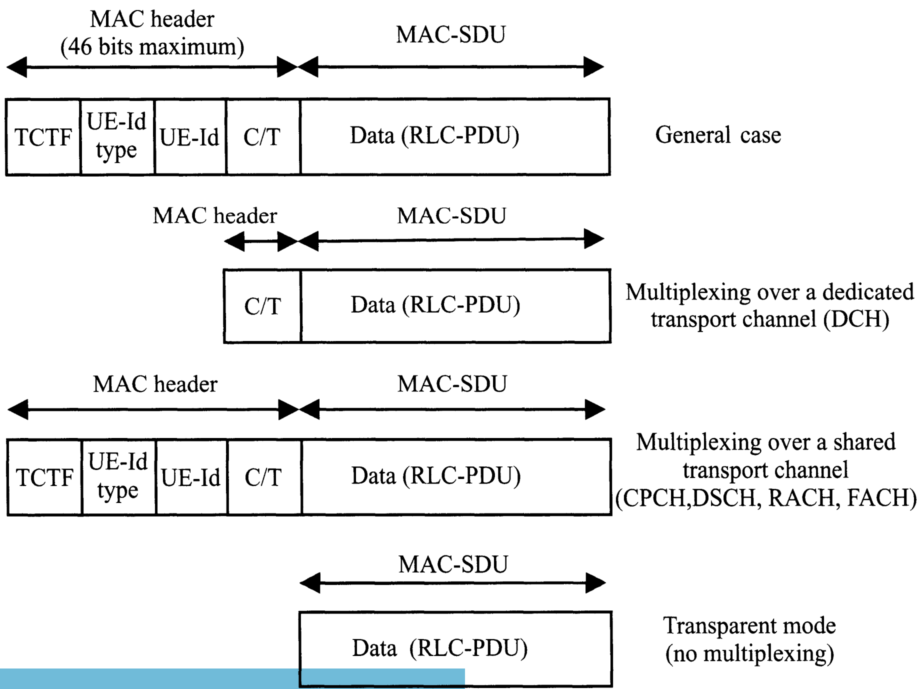


Figure 4.29 PDU format for the MAC layer.

The header part of the MAC-PDU is optional and its format is variable. When the MAC layer is used transparently, which is the case when no multiplexing is being carried out, the header is not present. The MAC-PDU is then equivalent to the MAC-SDU. This case occurs when a distinct transport channel is associated with each logical channel; one DCH then corresponds to each DCCH or DTCH.

In contrast, when the MAC layer is carrying out one of the multiplexing functions described above, the header is necessary to allow the receiving entity to de-multiplex the different logical channels. The header contains more or less information according to the type of transport channel used. Thus, when the user's logical channels are multiplexed on to a dedicated transport channel, the only information contained in the MAC header is the C/T field, which is a logical channel identifier. On receipt of the MAC-PDU, the remote entity can therefore tell to which logical channel it belongs.

When one or more of the user's logical channels are multiplexed on to a transport channel common to several users, the MAC header contains additional fields:

- the *Transport Channel Type Field* (TCTF), which identifies the type of transport channel used;
- the UE-Id, which contains the identity of the mobile receiving or sending the MAC-PDU. This field is used when several logical channels belonging to different users are multiplexed on to the same transport channel.

4.4.4 PDCP Layer

In version 99 of the UTRAN specifications, the function of the PDCP (*Packet Data Convergence Protocol*) layer is to compress the protocol headers of TCP/IP packets. This section shows why this compression is important for cellular networks and briefly describes the compression algorithm employed in the UTRAN.

Importance of TCP/IP Header Compression

Internet protocols were designed to be used on fixed networks with high traffic capacity and relatively low rates of transmission errors. They are therefore, in general, simple and robust, and use fairly large protocol headers.

Several studies of Internet traffic characteristics have shown that more than 40% of the IP packets that travel across the network are packets of very small size (40 bytes). These 40-byte packets, made up of 20 bytes of IP header followed by 20 bytes of TCP header, are control packets associated with the TCP connection and contain no user data.

The 552- and 576-byte packets correspond to the sizes of the PDUs frequently encountered on the Internet. The 1500-byte packets usually come from Ethernet networks linked to the public Internet (Figure 4.30).

Internet protocols, designed above all to be used on fixed networks without particular constraints in terms of bandwidth, are, in fact, unsuitable for cellular communication networks, whose capacity is much more limited and which are characterised by significant error rates.

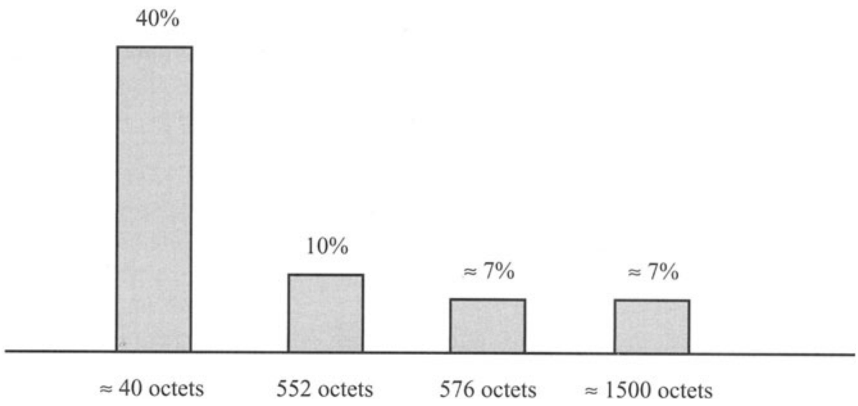


Figure 4.30 Histogram of the size of TCP/IP packets.

TCP is not the only transport protocol in the IP world. However, more than 90% of packets exchanged over the Internet are TCP packets, because they correspond to the most widely used applications, Web navigation and FTP transfers. The importance of an efficient mechanism for compressing the headers of IP datagrams transmitted over the radio interface is therefore clear. The radio resources will be used more efficiently and the number of erroneous packets will be reduced as a consequence of the reduction in the size of the packets transmitted.

Compression Algorithm for Protocol Headers

The PDCP layer of the UTRAN uses the TCP/IP header compression mechanism described in the following IETF specifications:

- RFC1144: *Compressing TCP/IP Headers for Low-Speed Serial Links*;
- RFC2507: *IP Header Compression.*

For the present these are the only mechanisms usable by the PDCP layer. However, PDCP has been specified in such a way as to allow other compression algorithms to be included in the future.

The TCP/IP compression algorithm specified in the IETF documents is based on the existence of a certain redundancy in the information contained in the headers. Between two consecutive TCP/IP packets, a very small number of data fields are going to change.

When IP packets are transmitted over the Internet fixed network, the header elements are, for the most part, used entirely for routing the IP datagrams across an interconnected network with a complex structure made up of different routing nodes. However, when the IP packets are transmitted across the radio interface, the routing information contained in each header becomes superfluous, because the communication between the mobile and the UTRAN uses a dedicated logical link (the DTCH).

Figure 4.31 shows in grey the TCP and IP header fields that are likely to remain unchanged throughout the duration of a TCP connection. Among these fields, there are, in fact, some that are supposed not to change during the life of a TCP connection:

- the IP protocol version;
- the IP addresses of the source and destination;
- the source and destination TCP ports.

Some other fields, such as the checksum, used for the detection of transmission errors, or the TCP sequence number, employed in the packet acknowledgement mechanism, vary from one packet to another and must therefore be transmitted to the remote entity.

Figure 4.32 shows the TCP/IP header after compression. The first byte identifies the TCP connection in use. The second byte is a field of bits indicating the presence or absence of elements situated further on in the header. The only data element that is kept identical is the TCP checksum, coded on 16 bits. The integrity control

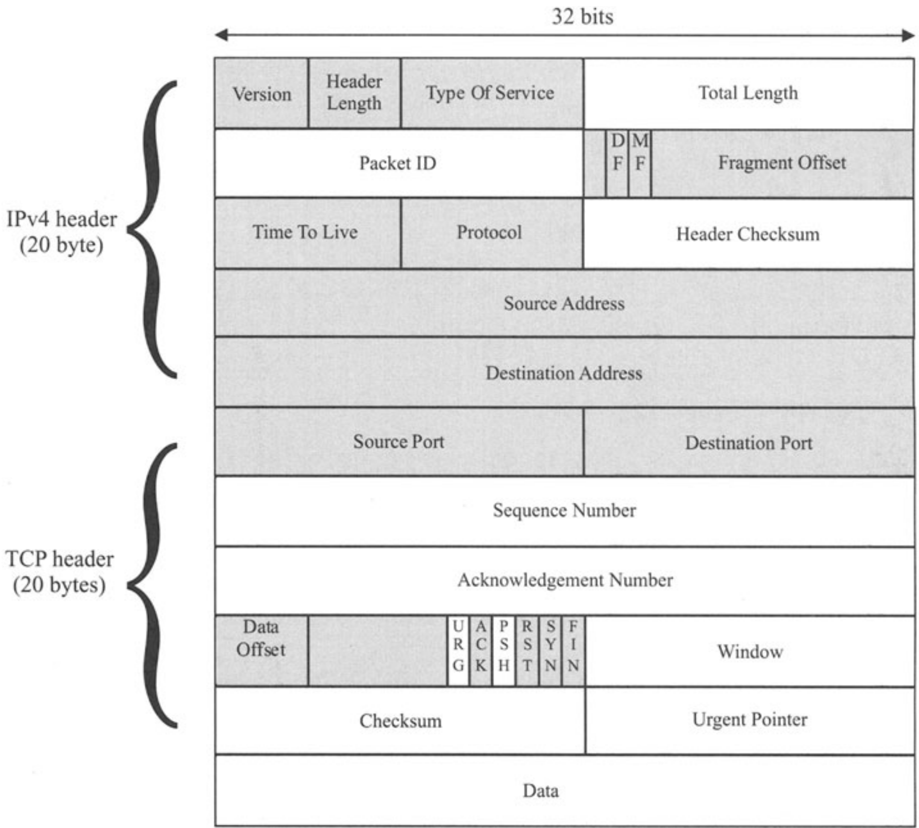


Figure 4.31 The TCP/IP header.

function of the TCP transport layer is thus unchanged. In the best case, therefore, the compressed header may only be four bytes long, corresponding to the non-grey areas in Figure 4.32.

To gain a little more space, certain fields, in particular fields W, A, S and I, use differential coding. Only the difference from the preceding value is transmitted, not the value itself. Thus the field containing the sequence number of the TCP header, coded on to 32 bits, occupies no more than 4 bits in the compressed header.

This mechanism presumes that the complete, uncompressed header is transmitted at least once. After that, it is possible to use the compressed header alone. The complete header is, however, used from time to time to refresh the context information held by the remote entity.

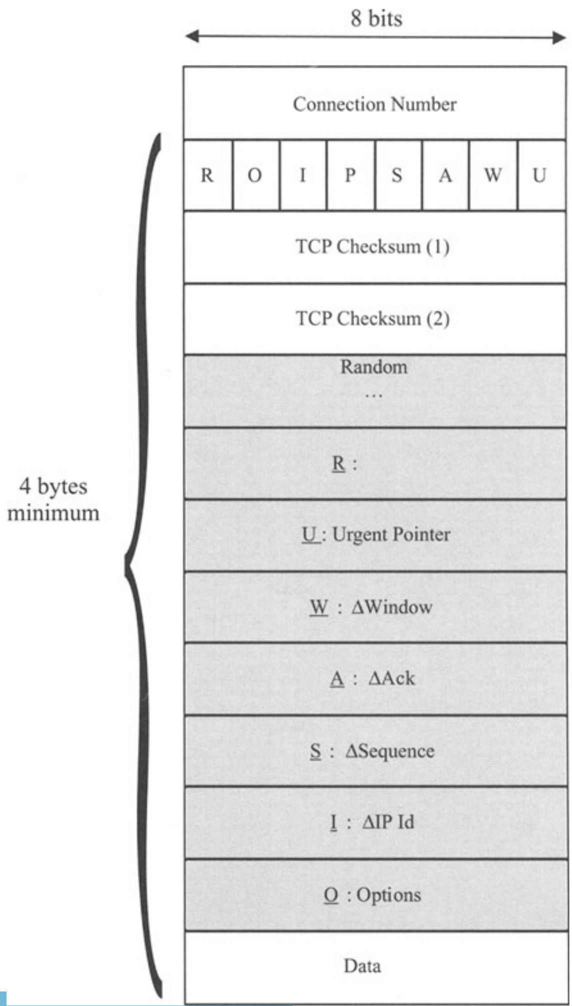


Figure 4.32 The compressed TCP/IP header.

The gain brought about by TCP/IP compression is quite significant. A TCP control packet will have its size reduced from 40 to 4 bytes. If we consider a TCP/IP packet of 512 bytes (the average size observed), the header alone represents 7.8% of the size of the packet. Thanks to the compression mechanism, this percentage is reduced to less than 1%. The gain is still more significant when IPv6 (version 6 of the IP protocol) is used, since its header alone is a minimum of 40 bytes.

4.4.5 BMC Layer

The purpose of the *Broadcast/Multicast Control* (BMC) layer is to provide for the distribution of user messages on the radio interface, to meet the needs of the service broadcast service. It also provides for repetition of information on the radio interface when this information needs to be broadcast several times.

4.4.6 General View

Figure 4.33 shows a general view of the protocol layers of the radio interface. The data packet received from the UTRAN core network is the N-PDU (Network PDU). In the case of an IP packet, the N-PDU header is compressed by the PDCP layer, that is to say, it is replaced by a smaller PDCP header. This new PDU is then segmented by the RLC layer, which adds its own header to each segment. The RLC-PDU is then handled by the MAC layer, which adds its own header if multiplexing is taking place.

This case is one of the most complex because it exercises many of the functions of the radio interface protocols. In the case of voice transmission, it is possible to imagine a much simpler method of working, in which the RLC and MAC layers would be used in transparent mode (no segmentation, no multiplexing of speech frames), the PDCP layer then being unnecessary.

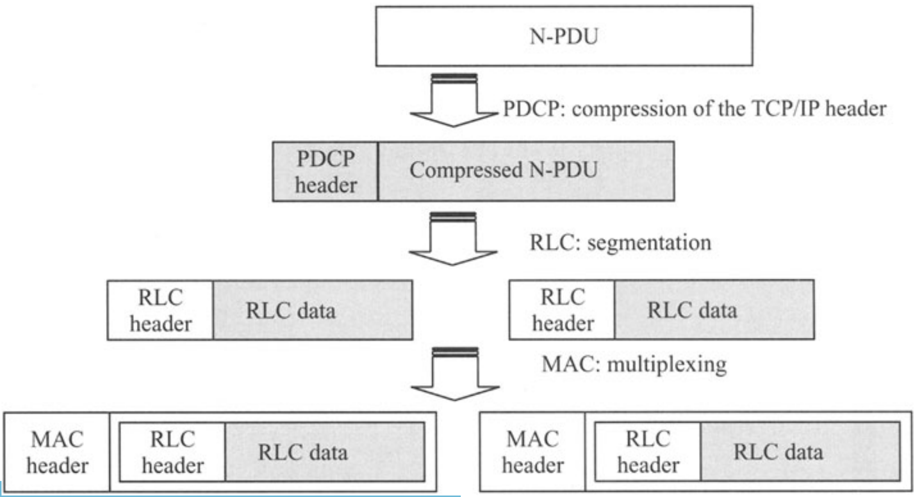


Figure 4.33 General view of the BMC layer.

4.5 Termination Points of the Radio Protocols

The GSM standard offers some flexibility with regard to the positioning of the radio protocols in the equipment of the BSS access network. For example, the precise position of the layers of the packet mode radio protocol (the RLC and MAC layers of GPRS) is never specified. Indeed, the standard is content to describe the different possibilities, leaving the final decision to the implementer. A logical entity, called the *Packet Control Unit* (PCU) and representing the point of termination of the radio interface protocols, is defined by the GSM specifications; it can be positioned at different points in the GSM access network. Figure 4.34 shows the different network architecture options. These options are not, of course, equivalent. The performance of the radio interface protocols and the size of the Abis and Gb interfaces vary from one option to another.

In the case of option 1, the PCU is included in the BTS and is just upstream of the channel coding unit. In options 2 and 3, the PCU is upstream of the BSC. In these two cases, the PCU can be either an independent piece of equipment or included in the BSC or the GSN.

In contrast to the GSM standard, the UMTS specifications define the functions of each piece of UTRAN equipment for each transport channel.

In the case of the DCH, radio protocol layers are all located in the SRNC (*Serving RNC*), i.e. the RNC that manages the RRC connection of the mobile to which the DCH is allocated. The functions of the physical layer are divided between the Node B and the SRNC, in order to support macro-diversity, a function that will be described later (Figure 4.35).

Between the Node B and the SRNC lies the Iub interface, or possibly Iub and Iur in the case that the Node B is not directly under the control of the SRNC.

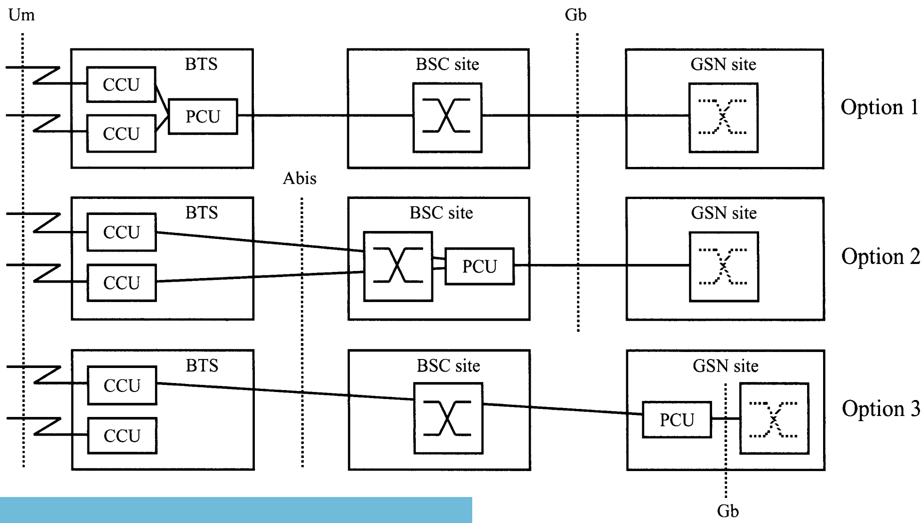


Figure 4.34 Implementation options for the GPRS radio protocols (source: 23.060).

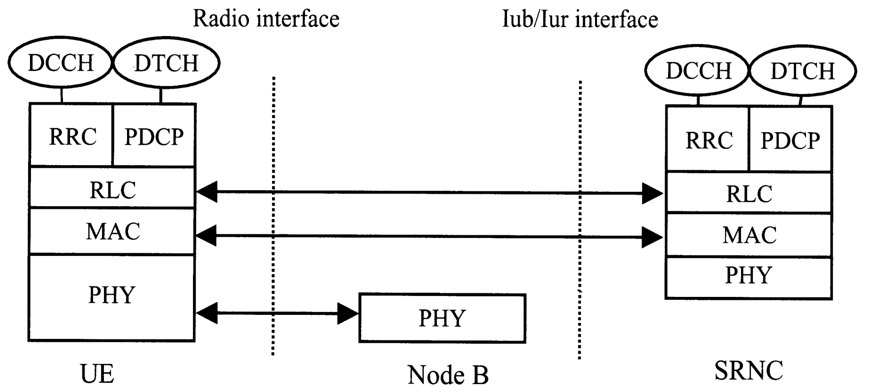
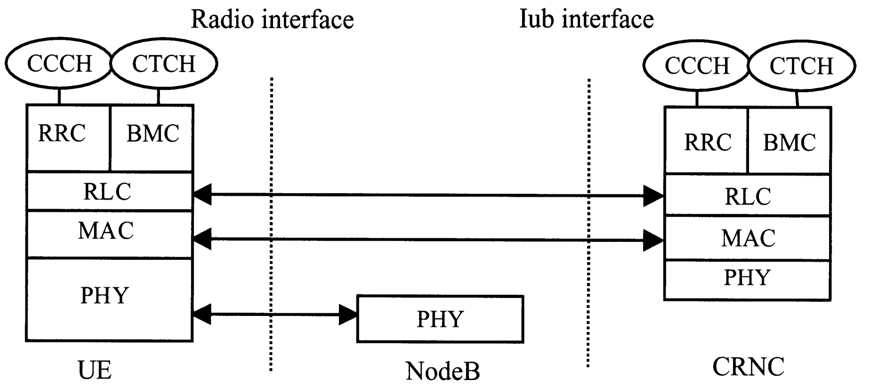


Figure 4.35 Termination points for DCH protocols.



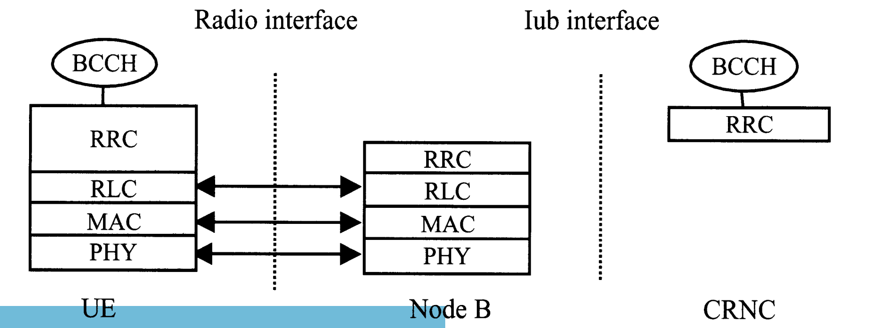


Figure 4.37 Termination points of protocols for the BCH transport channel.

In the case of certain logical channels supported by the FACH transport channel (for example, the CTCH or the CCCH), the protocol layers are necessarily situated in the CRNC (*Controlling RNC*), because the information carried by these logical channels is directly held by the CRNC (Figure 4.36).

The BCCH logical channel is used to broadcast control information into the UTRAN cells. This information is broadcast periodically on the radio interface by means of the RRC protocol, located in the CRNC corresponding to each cell. In order to avoid overloading the Iub interface by the repetition of these messages, the part of the RRC layer that handles the repetition is located in the Node B (Figure 4.37).

5 The Physical Layer of the Radio Interface

5.1 General Principles of CDMA

5.1.1 Multiple Access Techniques

In order to allow a group of mobile users to access the network simultaneously, it is necessary to share, in one way or another, the radio resources managed by the network operator. There are three main ways of providing such resource sharing or *multiple access*: frequency division multiple access (FDMA), time division multiple access (TDMA), and code division multiple access (CDMA).

Frequency Division Multiple Access

The principle of FDMA is to reserve for each user a portion of the available spectrum, which will be used throughout the call (Figure 5.1). Since FDMA is the oldest multiple access technique, it is used by numerous systems.

Time Division Multiple Access

Sharing by time, or TDMA, is an alternative to FDMA. The users of a TDMA system all use the same frequency band. The resource sharing is effected by allocating time slots to each user (Figure 5.2). To achieve full utilisation of the common resource the mobiles must transmit at very precise instances, which requires a periodic check of the transmission times, carried out by the network. In fact, depending on the distance between the mobile and the antenna of the network transmission

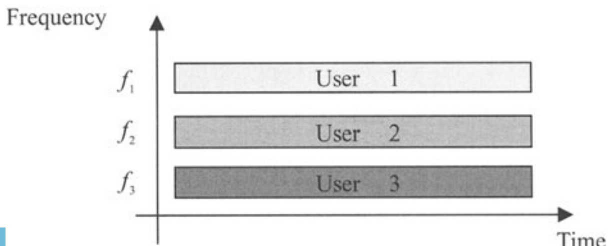


Figure 5.1 FDMA.

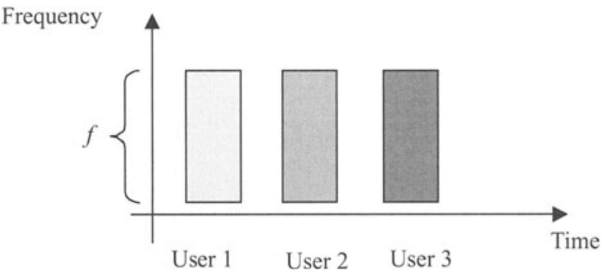


Figure 5.2 TDMA.

equipment, it is necessary regularly to correct the transmission time in order to take into account the propagation delay of the signal. This feature makes TDMA systems a little more complex than FDMA systems.

The radio interface of GSM networks uses a mixture of TDMA and FDMA (Figure 5.3). The GSM standard specifies 200 kHz carrier frequency bands, each being divided into eight time slots. The elementary resource allocated is thus characterised by the pair (frequency, time slot).

Code Division Multiple Access

CDMA is radically different from the two previous multiple access techniques. The users of a CDMA system all use the same band of frequencies at the same time, separation between the different users being achieved by a code identifying each one (Figure 5.4).

CDMA access can be compared to a group of people of different nationalities where everyone is speaking at the same time but using different languages. A new arrival who, for example, understood only English would be capable of extracting phrases in English from the surrounding hubbub, the conversations of the other participants appearing to be senseless noise. Of course, if the level of the general noise coming from other conversations becomes too great, it will be much more difficult, impossible even, for our visitor to understand what his compatriot is saying.

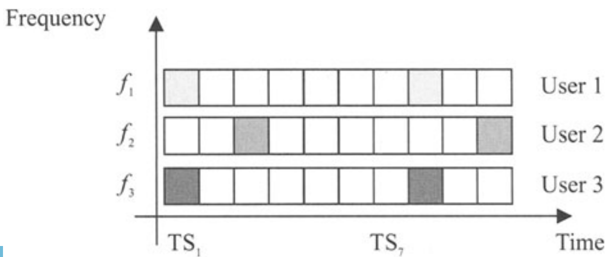


Figure 5.3 TDMA/FDMA.

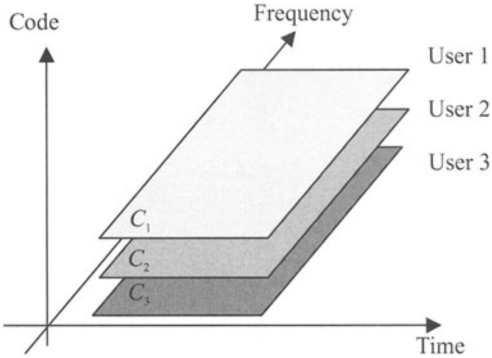


Figure 5.4 CDMA.

In an analogous manner, in a CDMA system, a certain number of signals are transmitted simultaneously in the same frequency band. They will all be received by the CDMA receiver, whose job it is to extract the data intended specifically for it from the set of signals received. This is possible because of the correlation properties of the codes used by the CDMA system. The receiver, knowing the code used by the transmitter, is able to recover the data transmitted. In contrast, the other signals, using different codes, will be eliminated because of their weak correlation with the code used by the receiver.

As in the case of our imaginary meeting, when the level of noise (or, rather, interference) is too high, the receiving entity will no longer be able to recover the signal transmitted. This phenomenon occurs when the limits of the system in terms of capacity are reached.

5.1.2 Correlation Properties

CDMA systems use codes endowed with special correlation properties. The codes are composed of fixed length bit strings.

The cross-correlation of two codes $S = (S_0, S_1, \ldots, S_N)$ and $T = (T_0, T_1, \ldots, T_N)$ of length N is defined to be:

$$R_{S,T}(i) = \sum_{j=0}^{N-1} (-1)^{S_i + T_{i+j}}$$

The autocorrelation of a sequence $S = (S_0, S_1, \ldots, S_N)$ is derived from the previous definition and is defined to be:

$$R_S(i) = \sum_{j=0}^{N-1} (-1)^{S_i + S_{i+j}}$$

In general, correlation and cross-correlation functions measure the degree of difference between two signals. In the case of code sequences, the result of this function is in fact a vector, whose ith component is equal to the number of positions in which S shifted i places left and T (not shifted) are equal less the number of places

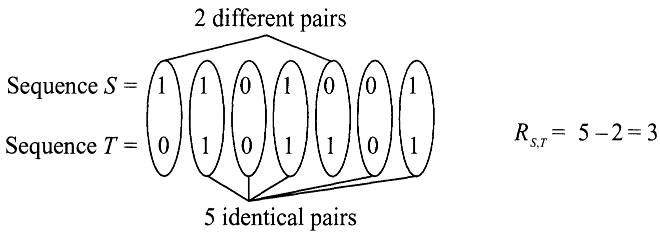


Figure 5.5 Example of the calculation of the correlation between two sequences.

in which they differ. Figure 5.5 shows an example of the calculation of the cross-correlation between two binary sequences. The value of the correlation function is in fact equal to the number of pairs that are identical minus the number of pairs that differ.

In a CDMA system, the codes used must satisfy the following correlation properties:

- the autocorrelation function $R_s(i)$ of each code is maximal when $i = 0$ and small or negative for other values of i;
- the cross-correlation between the codes is small or negative, or even zero in the case of a family of orthogonal codes.

Table 5.1 and Figure 5.6 show the value of the autocorrelation function for the particular code $S = (0111001)$.

Table 5.1 Values of the autocorrelation function of *S*.

i	*S(i)*	$R_s(i)$
0	0111001	7
1	1110010	−1
2	1100101	−1
3	1001011	−1
4	0010111	−1
5	0101110	−1
6	1011100	−1
7	0111001	7

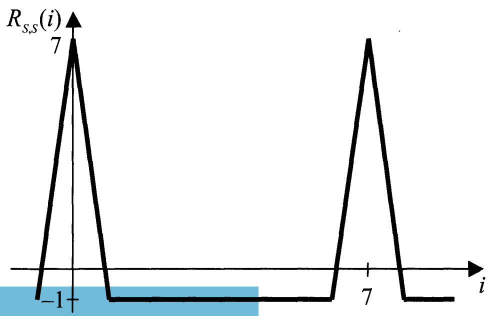


Figure 5.6 Autocorrelation of S.

Table 5.2 Values of the cross-correlation function of *S* and *T*.

i	*S(i)*	*T(i)*	$R_{S,T}(i)$
0	0111001	1101001	3
1	1110010	1101001	–1
2	1100101	1101001	3
3	1001011	1101001	3
4	0010111	1101001	–5
5	0101110	1101001	–1
6	1011100	1101001	–1
7	0111001	1101001	3

Note that the autocorrelation function reaches its maximum (7, the number of bits in the code) when $i \equiv 0 \bmod 7$[1] and takes the value –1 for all other values of *i*.

Table 5.2 and Figure 5.7 give the values of the cross-correlation function between the code *S*, used above, and the code *T* = (1101001). The codes *S* and *T* are not strictly orthogonal and hence the cross-correlation function is not equal to zero. However, it remains substantially below the maximum of the autocorrelation function.

5.1.3 Multiplexing and Demultiplexing Data

In CDMA, information to be transmitted is multiplexed simply by multiplying it by a code sequence allocated to the transmission. Figure 5.8 shows an example of a data flow $D(i)$ multiplied by the code $S(i)$ used in the previous paragraph.

In CDMA systems, the data rate of the coding bits is greater than that of the data bits (in our example, the data rate of the code sequence *S* is 7 times greater than that of the data bits). In the next section we will see that this has an effect on the bandwidth of the signal.

Before the data is multiplexed, the code bits and the information bits are encoded using the NRZ (Non-Return to Zero) convention. The data signal $D(t)$ is then simply multiplied by the code signal $S(t)$. The result $D(t) \cdot S(t)$ corresponds to the signal $S(t)$ modulated by the signal $D(t)$.

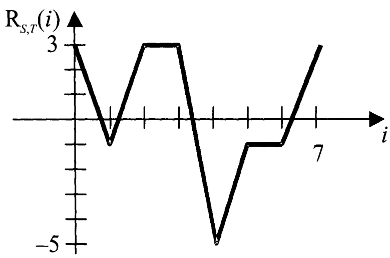


Figure 5.7 Cross-correlation of the codes S and T.

[1]The notation $m \equiv n \bmod k$, where m, n and k are integers (read as m is congruent to n modulo k) means that $(m-n)$ is exactly divisible by k.

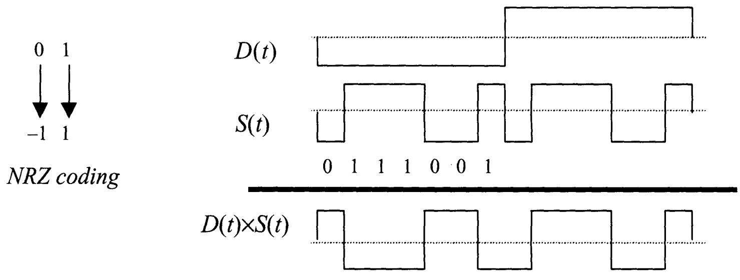


Figure 5.8 Transmission encoding.

At the receiver, the same operation is carried out. Assuming for simplicity that all the signals follow the same path, we can regard the signal received, $R(t)$, as identical to the signal transmitted, apart from a propagation delay of τ, so that $R(t) = D(t - \tau) \cdot S(t - \tau)$.

The signal received is then multiplied by $C(t - \tau)$, a copy of the code used by the transmitter, shifted by τ, to retrieve the data transmitted. The delay τ must be applied to the copy of the transmitter's code because, as we have seen, the maximum of the autocorrelation function is only attained when the two sequences are aligned (Figure 5.9).

A receiver using the code T will carry out the same operation but, because the codes S and T are not correlated, the result $D(t - \tau) \cdot S(t - \tau) \cdot C(t - \tau)$ will be rejected by the receiver.

In the case of simultaneous transmission of different signals, a composite signal, the sum of the different coded signals, is broadcast. This signal is therefore equal to

$$E(t) = D_1(t) \bullet S_1(t) + D_2(t) \bullet S_2(t) + \ldots + D_n(t) \bullet S_n(t)$$

where $D_i(t)$ is the signal to be sent to receiver i and $S_i(t)$ is the corresponding code sequence.

When the data received is de-multiplexed, the receiver τ will multiply the signal received by $S_j(t)$, to obtain

$$D_1(t - \tau) \bullet S_1(t - \tau) \bullet S_j(t - \tau) + \ldots + D_j(t - \tau) \bullet S_j(t - \tau) \bullet S_j(t - \tau) + \ldots \\ + D_n(t - \tau) \bullet S_n(t - \tau) \bullet S_j(t - \tau)$$

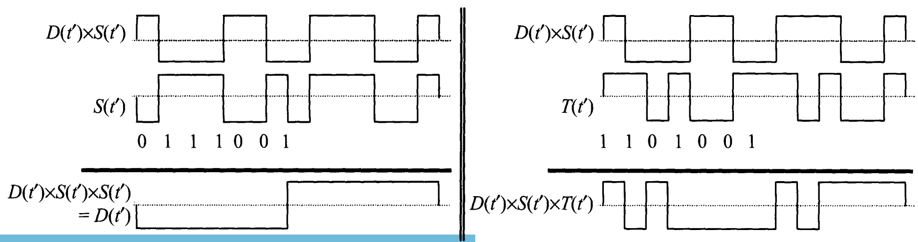


Figure 5.9 Decoding at the receiver (with $t - \tau = t'$).

For $i \neq j$, all the products $S_i(t - \tau) \cdot S_j(t - \tau)$ are close to zero and so the signal recovered by receiver j will be $D_j(t - \tau)$, which is what is required.

5.1.4 Use of Bandwidth

This section describes briefly the effect that the use of CDMA coding at the level of the UTRAN interface has on bandwidth requirements. The explanation offered is very simplified and its purpose is only to explain the principles that govern CDMA systems. There are numerous works on signal processing that address this subject in much greater detail.

The code sequences used by CDMA systems are of a pseudo-random nature and, for this reason, they are sometimes called pseudo-noise. It is this feature that enables us to calculate the bandwidth occupancy of the data signals.

Figure 5.10 shows the effect of the coding of a data signal in the frequency domain.

We have seen that, after NRZ encoding, a data signal $D(t)$ can be represented in the form of a series of impulses of amplitude +1 or −1, and of period T_d. The laws of signal processing show that the energy of such a signal is largely contained in the frequency band $\{-1/T_d, 1/T_d\}$. The maximum amplitude of $S(f)$, the power spectrum of the signal $D(t)$, is equal to T_d. (More generally, it is equal to a^2T_d, where a is the amplitude of the impulses.)

The period T_s of the bits used for the coding is less than Td. The frequency band of the coded signal, $\{-1/T_s, 1/T_s\}$, is therefore larger than that of the data signal. Similarly, the maximum amplitude of the power spectrum of the coded signal (proportional to T_s) is smaller than that of the data signal (proportional to T_d). The coding process thus spreads the power of the data signal over a broader frequency band. This type of technique is called *Direct Sequence Spread Spectrum* (DSSS), because the data signal is directly multiplied by the code sequence.

The ratio between the bandwidth of the coded signal and the bandwidth of the original signal data is known as the *spreading factor* (SF) or the *processing gain.* It is, in fact, equal to T_d/T_s, the ratio of the length of a data bit to the length of a coding bit. The concept of processing gain is important because it determines the capacity of CDMA communication systems. We will return to this point in Section 5.1.10.

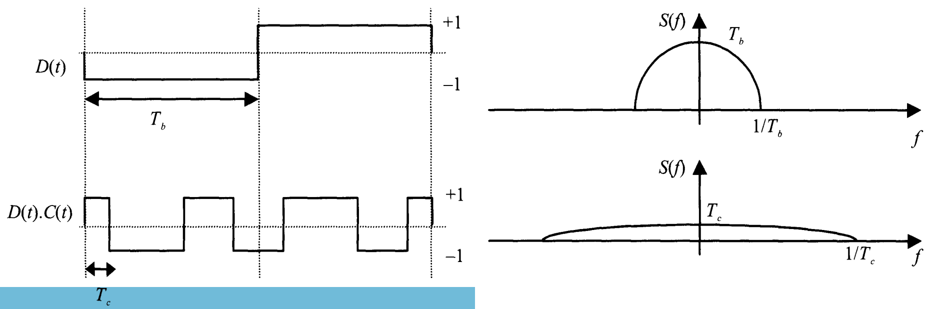


Figure 5.10 Spectrum of a CDMA signal.

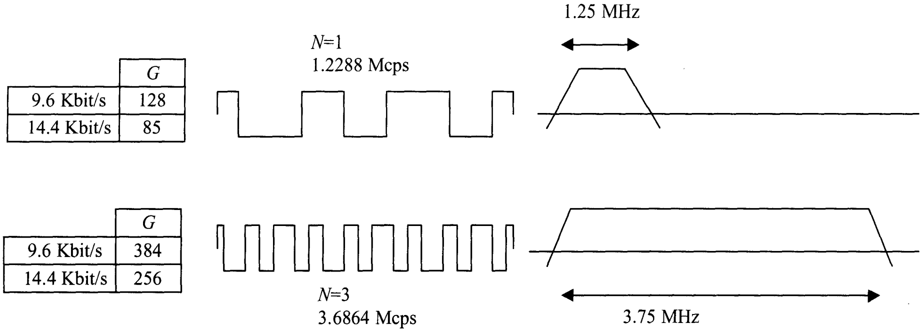


Figure 5.11 Example of spectrum spreading in cdma2000.

In the terminology of CDMA systems, the term *chip rate* (a *chip* being a code bit) is used to designate the bit rate of the coding sequence used. The second generation system IS-95, which makes use of CDMA technology, uses a chip rate of 1.2288 Mcps (1 Mcps = 1 Megachip per second = 10^6 chips per second). The UTRAN uses a higher chip rate of 3.84 Mcps. As a result the signal is spread over a much wider band than in IS-95. To distinguish between these two systems, the CDMA used in the UTRAN is known as wide band CDMA (WCDMA).

The chip rate used in the UTRAN is fixed. All the channels of the UTRAN radio interface are therefore spread in the same way, whatever their data rates.

The cdma2000 system uses a different technique. Several different chip rates can be used, all multiples of the basic chip rate used in IS-95: $N \times 1.2288$ Mcps, where N is one of 1, 3, 6, 9 or 12. Thus, for different user data rates, depending on the chip rate used, different spreading factors can be obtained, as Figure 5.11 shows.

Many other techniques for spectrum spreading exist, which are not employed in the UTRAN. By way of examples, we can mention the following:

- FH-SS (Frequency Hopping Spread Spectrum). In contrast to DSSS, the modulating frequency of the FH-SS is not constant but varies from one period to another.
- TH-SS (Time Hopping Spread Spectrum). The sender does not transmit continuously but only at certain precise moments. During the transmission periods, the whole frequency band is used.
- Hybrid techniques such as MC-CDMA (multi-carrier CDMA), used in the cdma2000 standard.

5.1.5 Multi-path Propagation (Rake Receiver)

In a multi-path propagation environment, the receiver receives different copies of the signal sent, shifted in time and corresponding to the different paths taken by the signal (Figure 5.12).

Two approaches are possible. In the first, only the signal coming from the dominant path is processed, the other copies being treated as interference and removed by the receiver. In the second approach, the different contributions received are

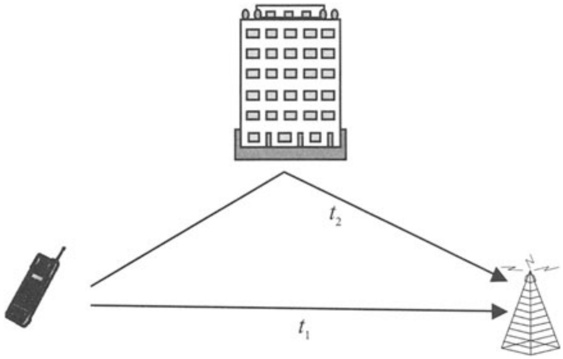


Figure 5.12 Multiple paths in an urban environment.

combined so as to benefit from the diversity of the transmission routes. The receiver used to implement this process is called a *rake receiver*. It can be used both on the uplink (in the Node B) and on the downlink (in the mobile). The principle is illustrated by Figure 5.13.

The different branches of the receiver correspond to the main transmission paths. In each branch, the signal received is correlated with a copy of the code used by the sender, shifted by the propagation delay for that path.

5.1.6 Codes Used in the UTRAN

OVSF Codes

The codes employed on the UTRAN radio interface are of the type known as *orthogonal variable spreading factor* (OVSF). In the UTRAN specifications, the OVSF codes are called *channelisation codes* or *channel codes*, because, thanks to the cross-correlation properties of OVSF, each sequence is specific to a single channel.

OVSF codes have the following properties.

- The sequences are strictly orthogonal, i.e. the cross-correlation between two code sequences is zero.

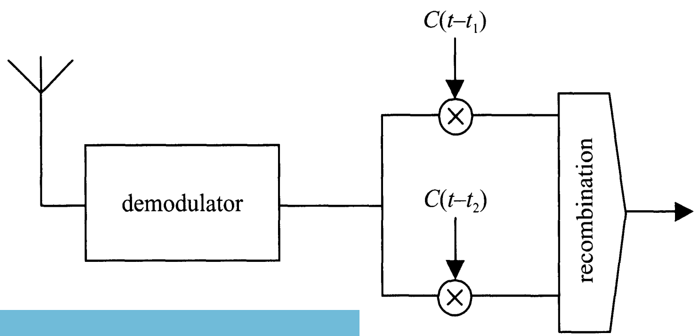


Figure 5.13 Example of a two-branch rake receiver.

- The sequences are not all of the same length, which allows for different spreading factors, depending on the data rate of the data to be transmitted.

We have seen that third generation systems offer a wide range of data rates, from a few kilobits to a few hundred kilobits per second. We have also seen, in the previous section, that the approach adopted by the cdma2000 radio interface was to offer a range of chip rates, thus allowing the system to offer different user data rates. In contrast, the UTRAN chip rate is fixed. The use of OVSF codes allows the SF to be varied, depending on the user data rate.

Figure 5.14 shows an example of variable SF resulting from the use of OVSF codes. In the first case, the user data rate is R. Each data bit is spread by a code sequence of length 2^n. In the second case, the user data rate is $2R$. To each data bit, there therefore correspond 2^{n-1} code chips, which corresponds to length of the code sequence used.

The OVSF codes used in the UTRAN are generated recursively, starting from the following definition:

$$C_1 = (C_{1,1}) = (1)$$

$$C_2 = \begin{pmatrix} C_{2,1} \\ C_{2,2} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$C_{2n} = \begin{bmatrix} C_{2n,1} \\ C_{2n,2} \\ C_{2n,3} \\ C_{2n,4} \\ \cdots \\ C_{2n,2n-1} \\ C_{2n,2n} \end{bmatrix} = \begin{bmatrix} (C_{n,1})(C_{n,1}) \\ (C_{n,1})(-C_{n,1}) \\ (C_{n,2})(C_{n,2}) \\ (C_{n,2})(-C_{n,2}) \\ \cdots \\ (C_{n,n})(C_{n,n}) \\ (C_{n,n})(-C_{n,n}) \end{bmatrix}, \; n \geq 2$$

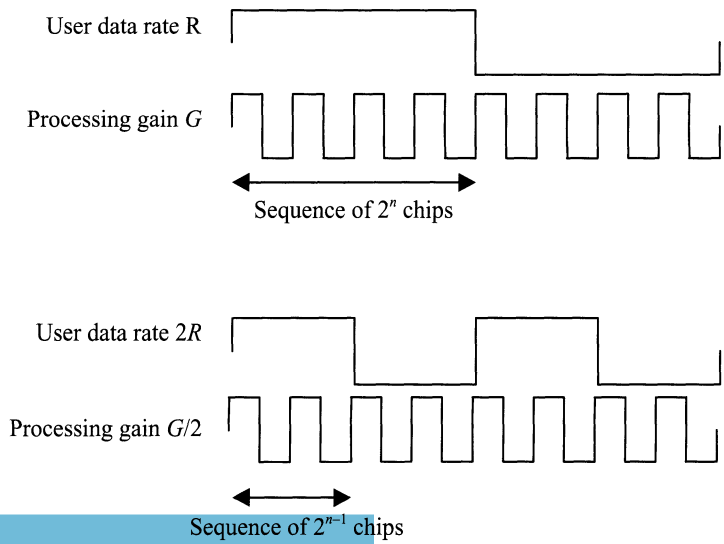


Figure 5.14 OVSF and variable processing gain.

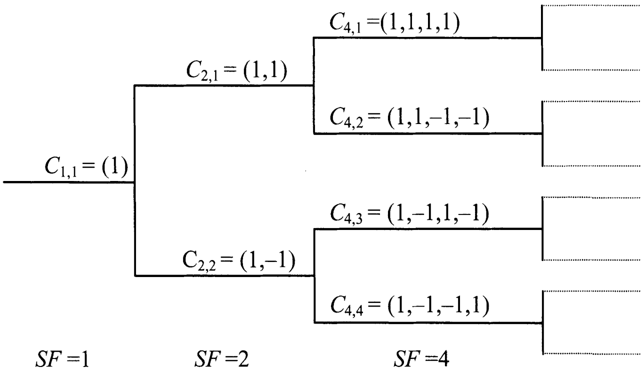


Figure 5.15 OVSF code tree (source: 25.213).

The OVSF codes can also be represented in the form of a tree, as shown in Figure 5.15. All the code sequences situated at the same level in the tree are of the same length and thus generate the same SF (an SF of 4 for all the codes $C_{4,i}$).

The tree structure brings out the cross-correlation properties of the OVSF codes. It is easy to see that two sequences at the same level of the tree are strictly orthogonal when they are aligned. For example, the cross-correlation of $C_{4,1}$ and $C_{4,3}$ is zero, according to the definition given in the previous section.

In contrast, two codes situated on the same branch of the tree, one being, for example, the child of the other, are not necessarily orthogonal. It follows that a code $C_{2n,i}$ can only be used if no other code is used that lies either on the sub-branches generated starting from $C_{2n,i}$ or on the path from $C_{2n,i}$ back to the root. This constraint is necessary to keep the codes used by the transmission system orthogonal. It seriously limits the number of codes that can be used simultaneously. Thus, when the four codes giving an SF of 4 (or the eight codes giving a gain of 8) have been used, no other code from the tree can be assigned. The maximum SF permitted by the UTRAN standard is 256 for uplink channels and 512 for downlink channels.

Scrambling Codes

The OVSF codes of the UTRAN cannot be used as they are because they do not give rise to pseudo-random sequences (for example, $C_{256,1}$ is a series of 256 ones). It is therefore necessary to resort to a second level of coding, this time using PN sequences, in order to achieve proper spectrum spreading. The codes used in this second level of coding are called *scrambling codes*.

There are a variety of methods of generating PN sequences. One of the simplest uses a generator based on shift registers (S) and binary half-adders (or, equivalently, exclusive OR gates), represented by the symbol $\oplus$.

Figure 5.16 shows an example of a PN sequence generator based on this principle. The initialisation sequence must not, of course, be zero since $0 \oplus 0 = 0$. The output of the generator is the sequence C of the preceding section.

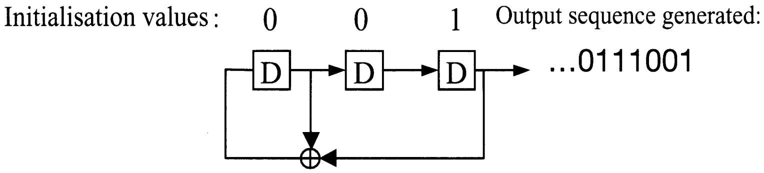


Figure 5.16 Example of a PN sequence generator.

Sequences generated in this way are called *linear feedback shift register* (LFSR) sequences. We note that they are periodic, of period $2^n - 1$, n being the number of shift registers in the generator.

The structure of the generator for the scrambling codes used in the UTRAN is a little more complicated. Figure 5.17 shows the structure of the generator for the codes used in the uplink direction. The generator has a double structure because the scrambling codes used by the UTRAN are complex codes, for which $C_{\mathrm{long},1,}n$ represents the real part and $C_{\mathrm{long},2,n}$ the complex part. We will consider the use of this complex code in the section on the operation of the physical layer.

Since the generator is made up of 25 shift registers, the code sequence generated consists of $2^{25} - 1$ bits. Because of the structure of the UTRAN radio channels, only 38,400 bits of the sequence are kept in the total sequence, in order to have available a scrambling code of a size equivalent to the size of a radio frame. (We will see, in the section on the structure of the physical channels, that the radio channels of the UTRAN are cut up into frames of 10 ms, which corresponds to 38,400 chips.)

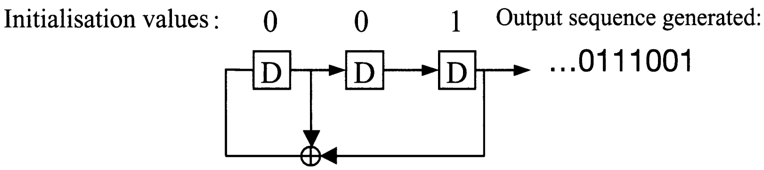


Figure 5.17 Scrambling code generator for the uplink (source: 25.213).

5.1.7 Use of Codes in the UTRAN

Allocation Principles

As a result of the way in which the OVSF codes are constructed, the cross-correlation properties between OVSF sequences are only guaranteed when they are synchronised. Figure 5.18 shows, in fact, how the sequences $C_{4,2}$ and $C_{4,4}$ are orthogonal when they are aligned. In contrast, when a shift is introduced between the two sequences, the cross-correlation properties are not necessarily satisfied. In the example in the figure, the two sequences seem perfectly correlated (the cross-correlation function is equal to 4) when the $C_{4,4}$ sequence is delayed by one bit.

	Aligned sequences		Sequences displaced by one place
$C_{4,2}$	1 1 −1 −1		1 1 −1 −1 1
$C_{4,4}$	1 −1 −1 1		1 1 −1 −1 1
Cross-correlation:	+ − + −	= 0	+ + + + = 4

Figure 5.18 Cross-correlation of two OVSF sequences.

In order to maintain the orthogonality between the transmission channels, the network must therefore synchronise the different codes used in sending signals from a single cell. The different channels will all be received by the mobile, with the same synchronisation (after compensation by the rake receiver in the case of multiple paths), allowing the mobile to decode correctly the signals intended for it.

In the uplink direction (i.e. mobile to network), it is impossible to obtain such synchronisation. One mobile simultaneously using several physical channels can synchronise the corresponding code sequences, but it is difficult to synchronise the code sequences for channels coming from different mobiles, because such control would impose too many constraints on the network.

Because of the synchronisation problems in the uplink direction, the code allocation principles are different for the two directions.

For the downlink direction, the network can use the whole of the OVSF tree in each cell, provided, of course, that the allocation constraints regarding parent and child in the same branch, described in the previous section, are respected. Further, each cell uses its own scrambling code C_s, guaranteeing independence between the channels of different cells (Figure 5.19).

For the uplink direction, orthogonality between the OVSF codes used by different mobiles cannot be guaranteed. Accordingly, the network must allocate to each

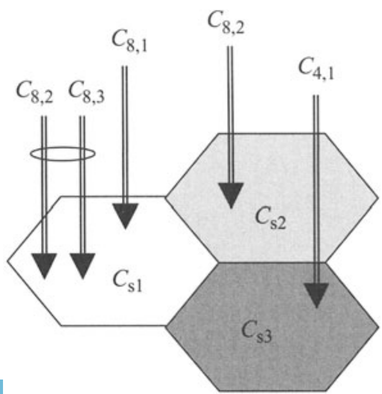


Figure 5.19 Example of code allocation for downlink channels.

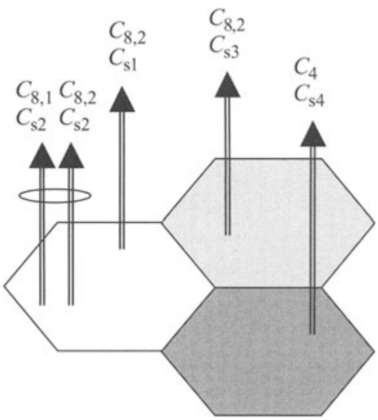


Figure 5.20 Example of code allocation for uplink channels.

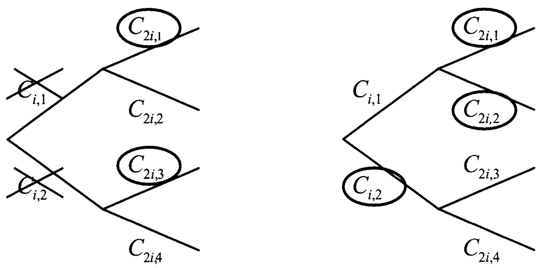
mobile its own scrambling code C_s. Each mobile can then use the whole of the OVSF tree. Figure 5.20 shows an example in which the code $C_{8,2}$ is used simultaneously by two mobiles in the same cell, but using two different scrambling codes.

Shortage of Codes

As a result of successive allocation and de-allocation of codes, it may happen that the allocation of new codes becomes impossible, even though the tree is only very sparsely filled. In the example in Figure 5.21, the codes $C_{2i,1}$ and $C_{2i,3}$ are used. In this case, the codes $C_{i,1}$ and $C_{i,2}$ are unusable because of the allocation rules for OVSF codes. In the case of a packed tree, it is possible to allocate the codes $C_{2i,1}$, $C_{2i,2}$ and $C_{i,2}$ simultaneously.

To alleviate the problem of shortage of OVSF codes, the UTRAN standard offers several possibilities.

The first method is to pack the codes used in such a way as to free as many of the branches of the tree as possible. This amounts to reconfiguring some physical



Codes $C_{i,j}$ unusable After 'packing' the tree

Figure 5.21 The effect of packing the OVSF tree.

channels whose code has changed. This procedure, entailing signalling exchanges between the network and the mobiles concerned, is carried out by the RRC layer.

The second method involves replacing a code with SF i by two codes with SF $2i$. In the example in Figure 5.21, this procedure amounts to dividing the data, which should have been transmitted on one channel using one code $C_{i,j}$, between two different channels using the codes $C_{2i,2}$ and $C_{2i,4}$.

The third method consists of using a second scrambling code in the cell. In our example, the codes $C_{2i,1}$ and $C_{2i,3}$ would be used with the scrambling code C_{s1}, and the code $C_{i,1}$ with the scrambling code C_{s2}, the independence of the corresponding channels being guaranteed by the cross-correlation properties of the codes C_{s1} and C_{s2}.

5.1.8 CDMA Planning

Planning the physical layout of the cells in GSM networks is a delicate operation. The local regulatory bodies (in the UK, the Radiocommunications Agency, now a part of Ofcom) allocate a set of frequencies to each network, the geographical distribution of the allocated spectrum then being the responsibility of the network operator.

Looked at in a very simplified way, cellular planning in GSM networks involves allocating a narrow band (200 kHz) carrier frequency to each cell. Because the multiple access technique used in the GSM radio interface is FDMA or TDMA, there must be a certain minimum distance between two cells using the same frequency in order to limit interference and collisions between the resources allocated to users. One classical procedure for cellular planning is to define a pattern, called the *reuse pattern*, which is repeated over all or part of the network. By way of example, Figure 5.22 shows the use of a pattern of size four.

There are some problems associated with this technique, in particular when the operator wants to improve the density of coverage by adding new cells to a region already covered. Such an operation cannot be carried out without modifying, more or less drastically, what is already in place.

As a general rule, the cells of a CDMA system all use the same frequency (Figure 5.23), a 5 MHz band in the case of the UTRAN FDD. Cellular planning for CDMA systems therefore seems much simpler than for systems using FDMA or TDMA. Indeed, this is one of the arguments most often put forward in favour of CDMA.

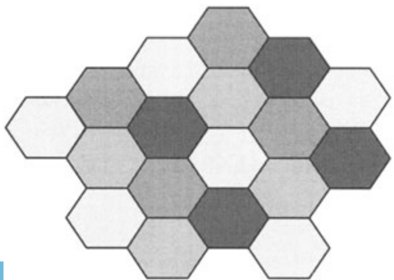


Figure 5.22 Cell planning in GSM.

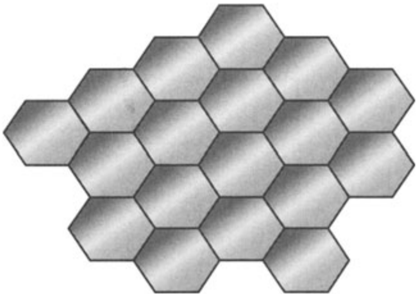


Figure 5.23 Cell planning in CDMA.

Despite the simplicity of the frequency configuration, CDMA networks are not without planning problems. We will see in the section on physical channels that two important parameters must be set for each cell:

- the secondary synchronisation code for the channel;
- the primary scrambling code.

There are only a limited number of these codes. It is therefore necessary to resort to reuse patterns, just like those used for frequency configuration in GSM. The task is, however, less complex because the UTRAN standard defines a substantial number of synchronisation and scrambling codes (64 secondary synchronisation codes and 512 primary scrambling codes are available).

5.1.9 Power Control

In a CDMA system, control of transmission power is particularly important, because users of the network, in contrast to users of a TDMA or CDMA network, are all using the same band of frequencies at the same time. Since each user can be considered a source of interference for the others, the capacity of a CDMA system is greatest when all the signals received by the receiver have the same power levels.

In the case of uplink transmission, in the absence of any control over the transmission power of the mobiles, those mobiles situated at the limit of the cell's range will be at a disadvantage in comparison with those close to the antenna of the base station, because they will be received more weakly. This is known as the near–far effect. It is therefore necessary to adjust the transmission power of the mobiles on the network. This is carried out through the TPC (*Transmit Power Control*) bits transmitted by the Node B on the dedicated physical channels, described at the end of this chapter.

In the downlink direction, all the signals are sent from a single source (the base station), which means that the transmission power levels can be balanced. Power control in the downlink direction is necessary, however, in order to limit interference generated in neighbouring cells.

5.1.10 Capacity Limits

In a transmission system, the correct receipt of information transmitted requires a certain minimum value for the ratio of the energy received per bit of information to the noise or level of ambient interference (Figure 5.24). This ratio is known as the *signal-to-noise ratio* (SNR) and can be expressed in the form E_b / N_o, where E_b is the energy per useful bit and N_o is the spectral density of the noise.

From the point of view of the CDMA receiver (R_i for user i), the useful energy per bit can be expressed in the form of the ratio between the power received, P_i for user i, and the useful data rate D, i.e. as P_i / D.

So far as the spectral density of the noise is concerned, the major source of noise in the system can be considered to come from the signals transmitted by the $N - 1$ other users of the system. If we assume that power control is perfect ($P_1 = P_2 = \ldots = P_N$), we can estimate the power of the noise as $(N - 1)\, P_i / B$, where B is the transmission bandwidth of the signal (3.84 MHz for UTRAN/FDD). The signal-to-noise ratio can then be written in the form:

$$\frac{E_b}{N_o} = \frac{\dfrac{P_i}{D}}{\dfrac{(N-1)P_i}{B}} = \frac{B}{D(N-1)}$$

If we neglect the difference between N and $N - 1$, we have

$$N \approx \left(\frac{B}{D}\right)\left(\frac{E_b}{N_o}\right)^{-1}$$

If we take a typical value of E_b/N_o to be 6 dB and D equal to 12.2 kbit/s (i.e. all the calls are voice calls), we obtain $N = 26 - 6 = 20$ dB, that is an estimate of 100 users per cell for telephone services. This result can also be shown graphically (Figure 5.25).

In a CDMA system, the resource that is shared is transmission power. In a much more critical way than in FDMA or TDMA systems, bad power allocation (for example because of poor or defective power control) will lead to a severe degradation of the network capacity.

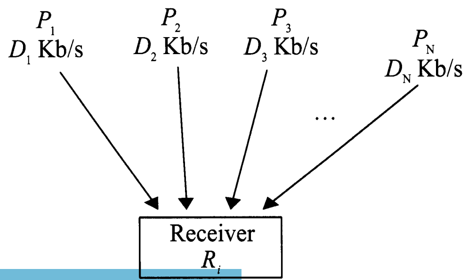


Figure 5.24 Signal-to-noise ratio.

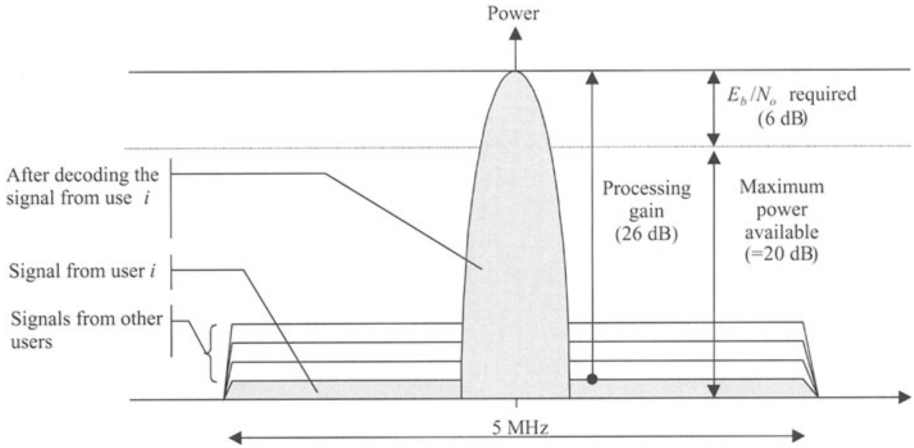


Figure 5.25 Graphical representation of the signal-to-noise ratio.

5.2 General View of the Physical Layer of the UTRAN

A certain number of operations are carried out by the physical layers of the fixed network and of the terminal. Figure 5.26 shows how these operations are linked in the case of a channel transmitting from the network to the mobile and vice versa. It will be seen that most of the operations are common to the up and down channels.

The operations carried out by the physical layer can be grouped into three sets:

- operations at the transport block level, carried out independently on the transport blocks of each transport channel;
- operations at the CCTrCH level, dealing with the functions common to the transport channels multiplexed on to a single CCTrCH;
- operations at the level of the physical channel.

Figure 5.26 also shows how the types of channel manipulated correspond to the different functions of the physical layer. The application of the transport format as required by the level 2 protocol layers is thus left entirely to the physical layer of the radio interface.

The remainder of this section describes briefly the functions used in the physical layer. Those functions needing more detailed explanations are described in separate sections.

Error Checking

The error-checking mechanisms include adding the CRC and channel coding. These mechanisms are described in later sections.

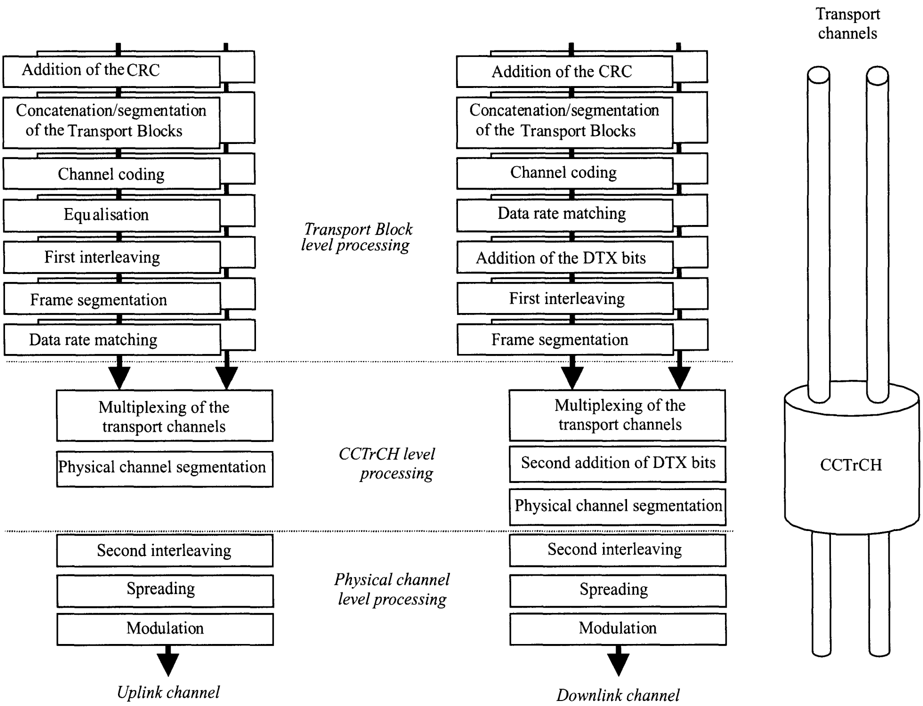


Figure 5.26 Operation of the physical layer.

Concatenation/Segmentation of Transport Blocks

The purpose of this operation is to prepare the data for channel coding.

When several transport blocks of a single transport channel are transmitted in the same time interval, the different blocks are concatenated into a single block which is then passed on to the channel-coding function. However, in order to limit the complexity of this function, this block is segmented if its length exceeds a certain value (the maximum size of a data block to be coded is 504 bits for convolutional coding and 5114 bits in the case of turbocode).

Equalisation and Data Rate Matching

The purpose of the data rate matching function, described in a later section, is to adjust the size of the blocks coming out of the channel coding to the capacity of the physical channel, the number of bits in a coded block not necessarily being equal to the number of bits that a physical channel can carry. The equalisation function has a rather similar role, since, by adding bits, it allows the length of the coded block to be adjusted to match the segmentation into frames.

As an example, consider a transport block whose transmission time is 40 ms, which will be cut into four segments of equivalent total length, each one transmitted in a 10 ms frame. The purpose of the equalisation function is to ensure that the size of the coded transport block is indeed a multiple of 4, by adding the necessary number of bits.

First and Second Interleavings

The purpose of the interleaving functions is to optimise the error control processes. The functions are described in later sections.

Frame Segmentation

The purpose of this function is to segment the coded transport blocks according to the TTI (transmission time interval) of the transport channel. For example, a transport block on a channel with TTI equal to 20 ms will be cut into two segments, each segment being carried by a different radio frame.

Multiplexing of Transport Channels

This function is used when different transport channels are multiplexed on to one CCTrCH. The segments of transport blocks from each of the multiplexed transport channels are simply concatenated.

Physical Channel Segmentation

This function is employed when several physical channels are used to transport the data from a particular CCTrCH.

Spreading and Modulation

The purpose of these operations, described in later sections, is to adapt binary data for transmission over the radio interface.

5.2.1 Error Control

In order to protect data against transmission errors on the radio interface, special protection methods are included in the physical layer of the UTRAN.

The UMTS specifications offer a wide choice of error control methods. The choice of the method to be used is carried out by the UTRAN, taking into account the quality of service requirements placed on each radio access bearer by the core network.

In a general way, error control methods can be divided into two groups:

- Techniques of the ARQ (Automatic Repeat Request) type. When an erroneous frame is received, the receiver can either reject the frame or demand its retransmission. This function is typically provided by the RLC layer of the UTRAN.
- Forward error correction (FEC) techniques. In this case, special coding is applied to the data to be transmitted. This operation, called channel coding,

generates extra bits that will be used by the receiving entity to correct the transmission errors.

The rest of this section describes briefly the methods of error correction used by the UTRAN physical layer.

Error Detection Using CRCs

Error detection using a cyclic redundancy check (CRC) is a mechanism widely employed in transmission networks. In particular, it is used in the radio interface of GSM networks (Table 5.3).

The specifications of the UTRAN define five layers of error detection, corresponding to five different sizes of the CRC, going from 0 (no CRC error detection) up to 24. The choice of which level of error detection to apply is made by the UTRAN, on the basis of the quality of service attributes of the corresponding RAB.

Convolutional Codes

The plan of a convolutional coder is comparatively straightforward since it is made up of shift registers and binary half-adders (otherwise known as exclusive OR gates), represented by the ⊕ sign. Convolutional codes are also used in GSM networks.

The first coder shown in Figure 5.27 produces two output bits for each input bit. The *code rate*, the ratio of the number of input bits to the number of output bits, is thus ½, because a sequence of n bits will be encoded on to $2n$ bits. An 8-bit header is added to the n bits of information to be coded. This provides protection equivalent to that obtained by the bits placed at the end of the sequence.

The coder with a code rate of ⅓ follows the same principle. This coder produces more robust protection than the previous one, at the cost of higher redundancy (a sequence of n bits will be coded on to $3n$ bits). Decoding such a sequence in the receiving entity is a much more complex operation than coding it. The Viterbi algorithm is used.

Turbocodes

Turbocodes are made up of convolutional coders used in parallel. Turbocodes were introduced into the standard because of their error correction performance.

Table 5.3 The polynomials used for error detection.

Polynomial used	CRC size
$g_{CRC24}(D) = D^{24} + D^{23} + D^6 + D^5 + D + 1$	24
$g_{CRC16}(D) = D^{16} + D^{12} + D^5 + 1$	16
$g_{CRC12}(D) = D^{12} + D^{11} + D^3 + D^2 + D + 1$	12
$g_{CRC8}(D) = D^8 + D^7 + D^4 + D^3 + D + 1$	8
–	0

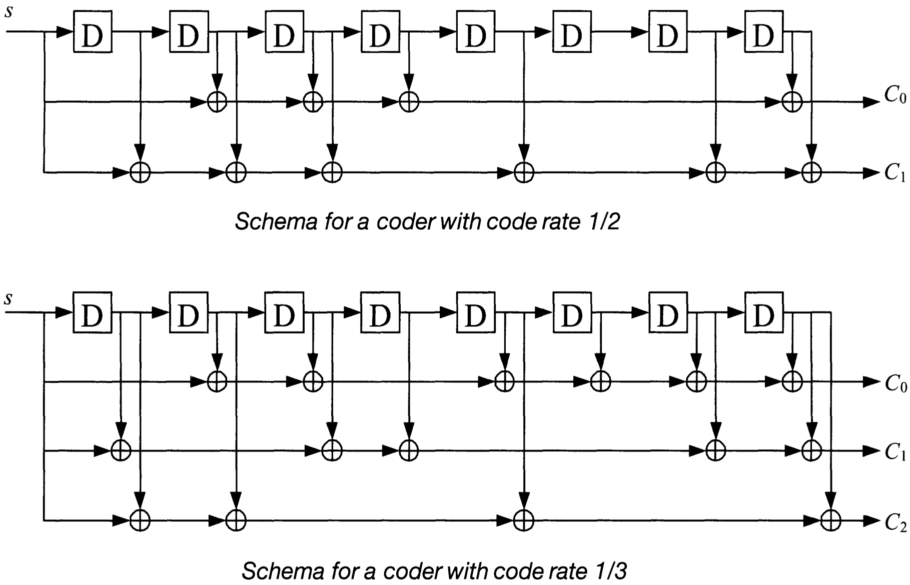


Figure 5.27 Convolutional coding in the UTRAN (source: 25.212).

By way of information, Figure 5.28 shows the schema of the turbocoder used in the UTRAN. The code rate is ⅓.

As in the case of the CRC, the type of channel coding (convolutional coding or turbocode) is chosen by the UTRAN on the basis of the quality of service attributes negotiated between the user and the core network.

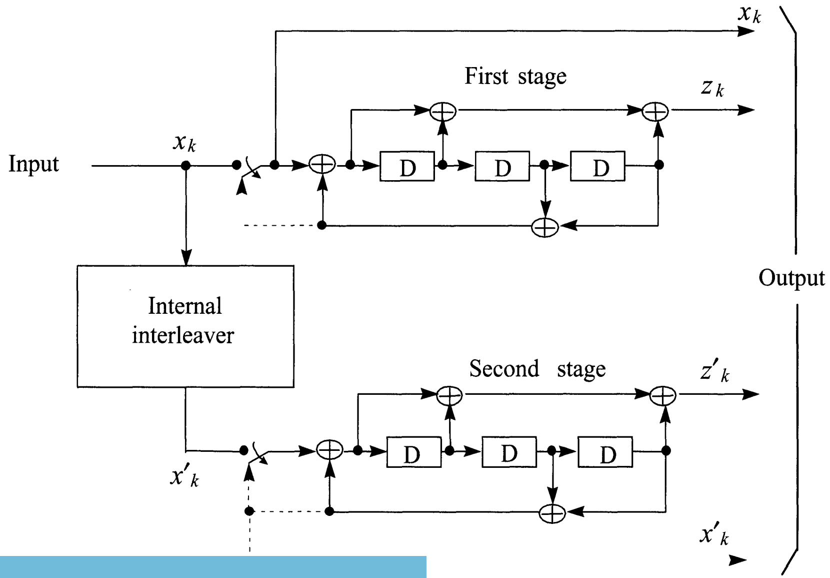


Figure 5.28 Schema of the UTRAN turbocoder (source: 25.212).

5.2.2 Rate Matching

The purpose of this operation, carried out after the channel coding phase, is to adjust the size of the coded blocks to the parameters of the physical channels of the radio interface. These physical channels in fact offer a fixed number of bits per frame, which does not necessarily match the size of the data packets supplied by the higher layers.

When radio resource is being allocated to the user, the network must choose a code whose spreading factor is as close as possible to the data rate requested. The data rate of the chosen resource may therefore be greater or less than the rate of the data to be transmitted. The network must work out the best compromise, knowing that allocating too much resource risks jeopardising the capacity of the network as a whole.

When the size of the blocks provided by the channel coding function is greater than that of a physical block (depending on the maximum number of data bits that a radio frame can contain), certain bits of the coded block are suppressed. This is known as 'puncturing'. Puncturing is based on an algorithm for determining which bits can be suppressed, i.e. the bits whose suppression will not damage the error control too much.

On the other hand, when the size of the coded blocks is less than that of the physical blocks, certain bits are repeated in order to fill the frame to be transmitted.

The data rate matching operation is carried out at the level of each transport channel. It must, however, take into account the respective sizes of the transport blocks that will later be multiplexed on to the same CCTrCH.

5.2.3 Interleaving

The purpose of error control methods used on the UTRAN radio interface, such as convolutional coding, is to protect the data transmitted against the high error rates associated with radio transmission. The ability that these methods have to detect and correct errors is reduced when the errors occur in groups, which is unfortunately the case with errors due to radio transmission. In order to make the distribution of errors more random and thus to improve the error correction performance, a phase in which the bits transmitted are mixed is added to the physical layer processing chain. We saw in the general description of the physical layer of the UTRAN that there are two levels of interleaving: in the transport block and at the level of the radio frame.

Interleaving at Transport Block Level

The first level of interleaving is carried out on the transport block, after channel coding has been applied. The bits of the coded transport block are interleaved before cutting the block into segments, which will then be carried by different radio frames. If an error occurs in a frame, the erroneous bits will in fact be distributed over the whole of the transport block after the bits have been put back into order by the receiver (Figure 5.29).

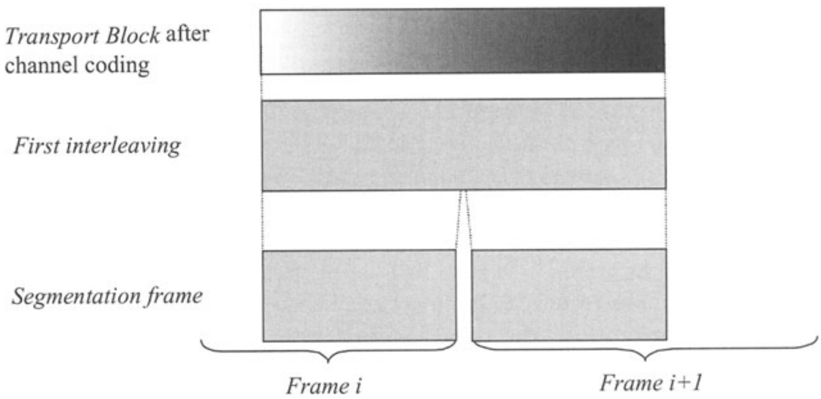


Figure 5.29 First interleaving.

Interleaving at Frame Level

The second level of interleaving takes place on the radio frame. The bits of the radio frame, made up of one or more transport block segments, are interleaved before cutting the frame into 15 slots. If one of the slots is corrupted during transmission, the operation of putting the bits back into order will distribute the erroneous bits over the whole of the radio frame. The errors will thus be distributed over the different segments of the transport blocks making up the frame (Figure 5.30).

Interleaving Process

The process used in the UTRAN for the two levels of interleaving is a simple mechanism called *block interleaving*, shown in Figure 5.31. The data bits taken

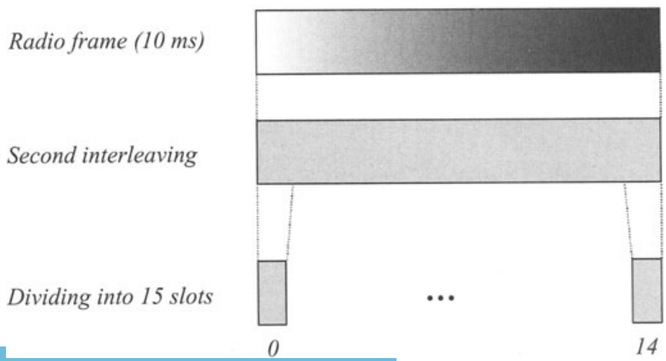
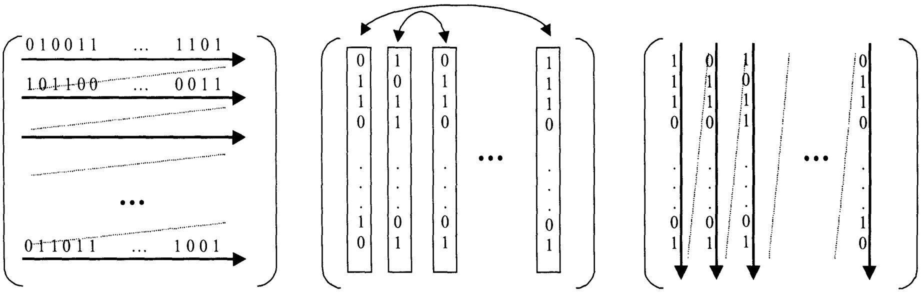


Figure 5.30 Second interleaving.



1 - Arranging the bits by rows *2 - Permutation of the columns* *3 - Reading the matrix by columns*

Figure 5.31 The three phases of interleaving.

sequentially are placed into the rows of a matrix. The number of columns of the matrix depends on the transmission time interval (TTI) for the first interleaving and is fixed at 30 for the second interleaving. The functions of data rate equalisation (for the uplink channels) and data rate matching (for the downlink channels), previously carried out, make it possible to ensure that the number of bits to be interleaved is divisible by the number of columns, so that the matrix is completely filled.

The columns of the matrix thus formed are permuted in a way specified in the standard and which depends on the number of columns in the matrix. The matrix is then read column by column.

The value of the TTI associated with the transport channel plays an important role in the effectiveness of the interleaving mechanism.

In the case of a transport channel with a low TTI (10 or 20 ms, for example), the benefit of the interleaving is reduced because the possible transmission errors remain relatively grouped. Each coded transport block will, in fact, be distributed over only a very small number of radio frames (one or two in our example).

In contrast, in the case of a high TTI (80 ms, for example) each coded transport block will be carried by eight frames. The use of a high TTI coupled with the two levels of interleaving allows the erroneous bits in a radio frame to be distributed over a large number of coded transport blocks, thus improving the effectiveness of the error control mechanisms.

The use of high TTI values entails a transmission delay equal to the length of the TTI, which can be prejudicial to the quality of service provided to the user. These high values are therefore kept for services that are sensitive to transmission errors and not subject to real-time constraints.

In the contrary case, voice telephony for example, the UTRAN must use low values of the TTI. In the case of voice telephony, transmission delays greater than 100 ms quickly become irritating to users.

5.2.4 Spreading and Modulation

This section describes the spreading and modulation operations carried out by the physical layer before passing the data to the radio interface.

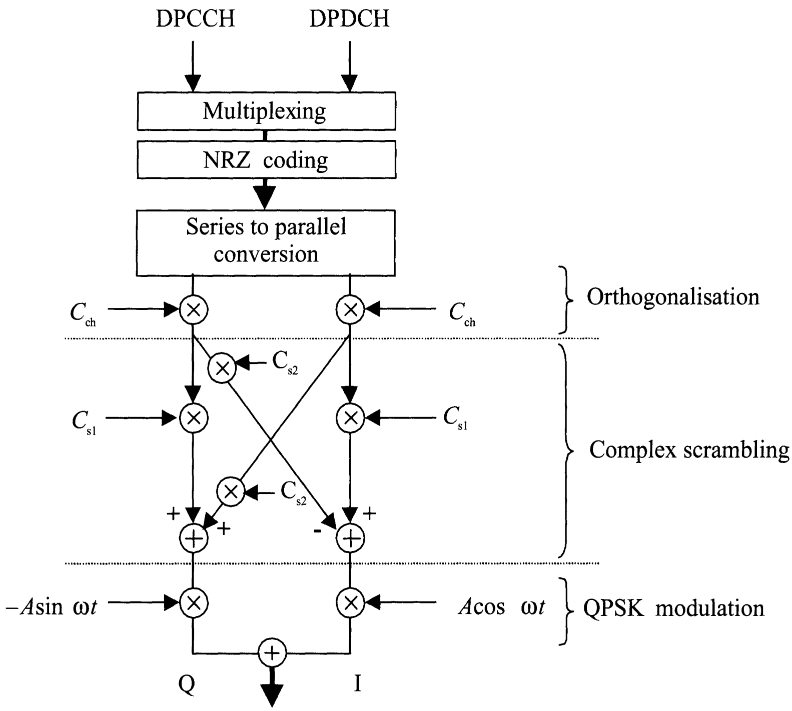


Figure 5.32 Processing the downlink physical channel DPCH.

Figure 5.32 shows the processing sequence for a downlink physical data channel DPCH. The bits supplied by the higher levels (dedicated physical data channels) are time multiplexed with the physical layer control bits (from the dedicated physical control channels). The resulting flow of bits is NRZ coded, then separated, the odd-numbered bits being directed towards the I branch and the even-numbered bits to the Q branch.

The bits on the two branches are encoded with the code C_{ch} of the OVSF tree, then scrambled with the complex scrambling code $C_s = C_{s1} + jC_{s2}$. The complex scrambler carries out a multiplication between a complex signal $S_I + jS_Q$ and the complex code $C_s = C_{s1} + jC_{s2}$, the result being a signal with a data rate of 3.84×10^6 complex symbols per second. QPSK (quadrature phase-shift keying) is then applied before passing the signal to the radio interface.

For the physical channel DPCH in the uplink direction, the situation is slightly different. The operations carried out are almost identical except that the control bits are not multiplexed with the data to be transmitted but are passed directly to the Q branch (Figure 5.33). We will examine the reason for this difference when we describe the structure of the physical channels.

QPSK Modulation

The purpose of this modulation is to adapt the transmission infrastructure to handle a digital signal made up of binary elements.

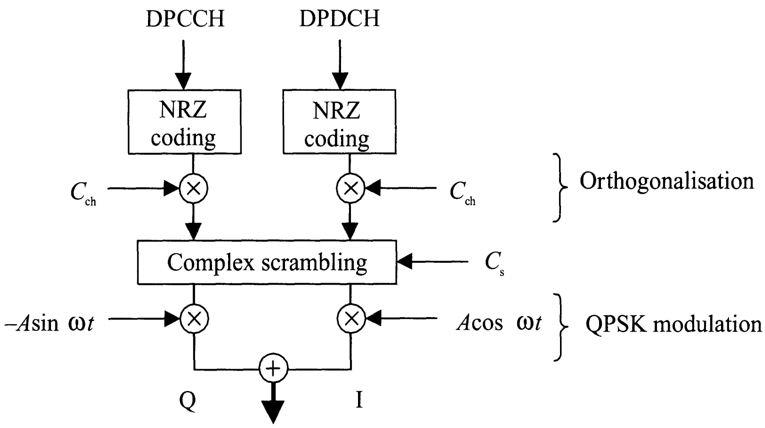


Figure 5.33 Processing the uplink physical channel DPCH.

In general, a modulated signal can be represented in the following form:

$$S(t) = A(t) \cdot \cos(2\pi f_c t + \varphi(t))$$

According as to whether the useful information is carried by the amplitude $A(t)$ or by the phase $\varphi(t)$, we speak of *amplitude modulation* or *phase modulation*. Certain types of modulation, such as quadrature amplitude modulation, combine the two techniques.

QPSK is a member of the family of phase modulation techniques. A simplified schema for a QPSK modulator is shown Figure 5.34. It is made up of two branches described as 'in quadrature', because they are modulated by carrier frequencies that are $\pi/2$ out of phase.

The NRZ coded bits are distributed on to the I and Q branches of the modulator, to be finally modulated by a carrier of frequency f_c. Multiplication of the carrier frequency by the NRZ symbols has the effect of changing the phase of the carrier by π at each change of symbol (Figure 5.35).

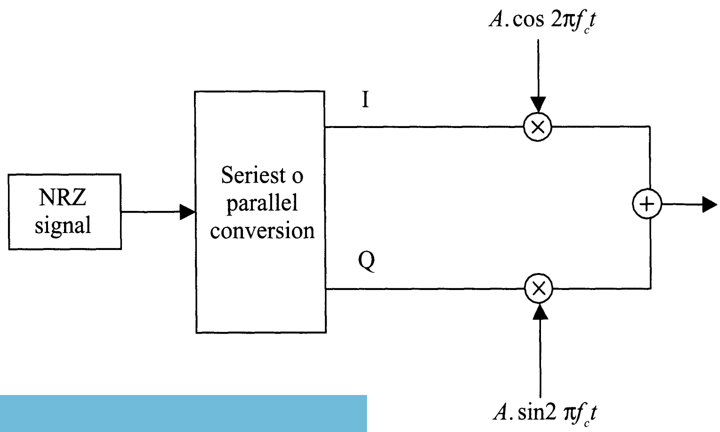


Figure 5.34 Simplified view of a QPSK modulator.

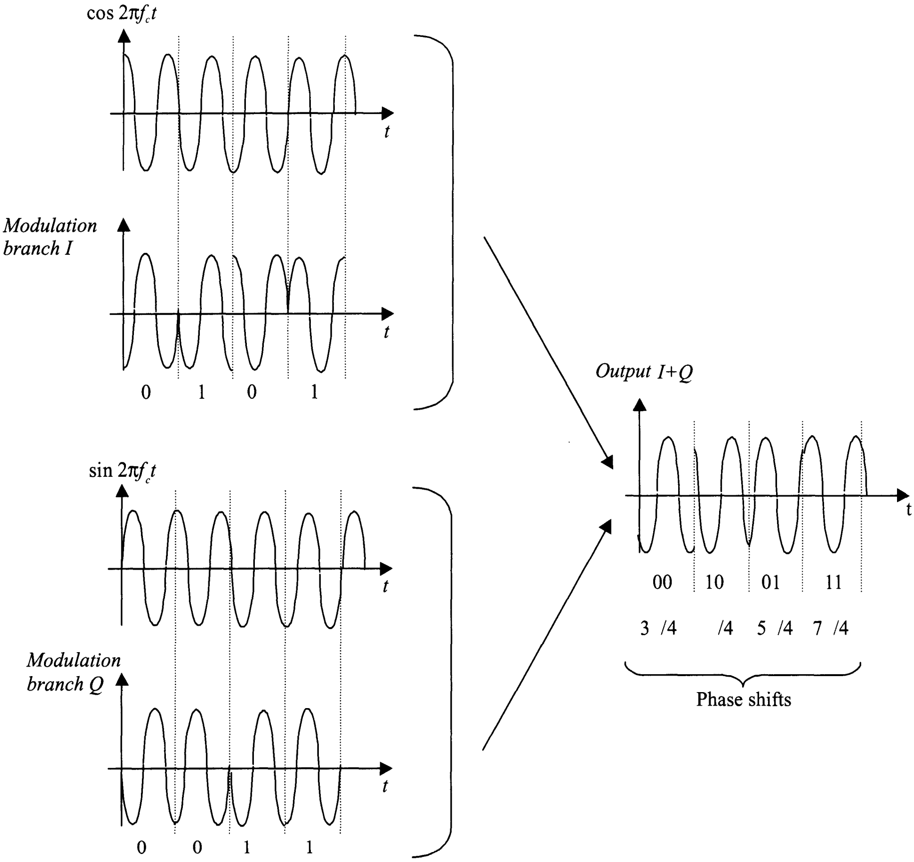


Figure 5.35 QPSK modulation.

The signal emitted by the modulator is the result of summing modulated carriers from the I and Q branches, which is, in effect, a carrier of frequency f_c with phase shifts of $n\pi/4$, n being equal to 1, 3, 5 or 7, depending on the combination of symbols on the I and Q branches.

Figure 5.36 shows the phase (or constellation) diagram for QPSK modulation. Since each QPSK signalling event carries two bits of information, the diagram contains four states, corresponding to the four phase differences with respect to an unmodulated reference carrier. QPSK demodulation requires the receiver to generate two synchronised carriers $\cos 2\pi f_c t$ and $\sin 2\pi f_c t$ to perform coherent demodulation of the received signal. To avoid a potentially large phase shift, the I and Q branches are time-shifted by half a bit, on the uplink only.

5.3 Structure of the Physical Channels

This section describes the structure of the principal physical channels of the UTRAN. In order to better understand their purpose, each physical channel is described in terms of the basic functions used by mobiles and by the fixed network:

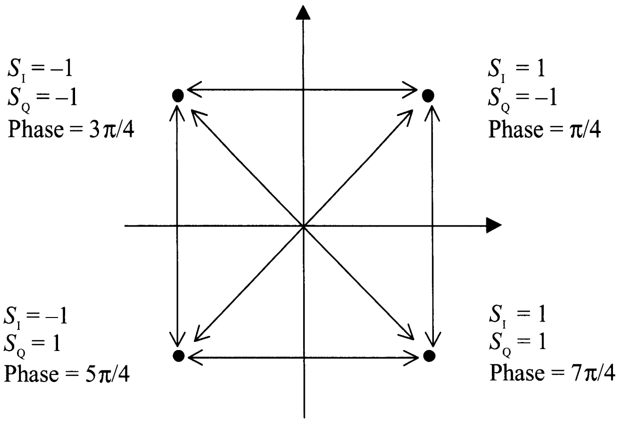


Figure 5.36 Phase diagram for QPSK modulation.

- initial cell search (CPICH, P-SCH and S-SCH);
- reading system information (P-CCPCH);
- broadcasting paging messages (S-CCPCH);
- sending the initial message when a call is being set up (RACH and AICH);
- exchange of data on a dedicated channel (DPCH).

Transmission over the UTRAN radio interface is split up into frames of length 10 ms, each frame itself being segmented into 15 slots. The UTRAN chip rate is 3.84 Mcps so each slot potentially contains 2560 code elements (3.84 Mcps $\times$ 10 ms/ 15 slots = 2560 chips) (Figure 5.37).

This frame structure is applicable to almost all the physical channels of the UTRAN, both uplink and downlink.

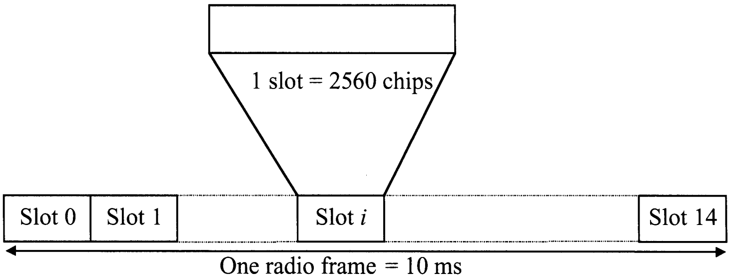


Figure 5.37 Frame structure of the radio interface.

5.3.1 Initial Cell Search

When it is switched on, the mobile carries out a number of operations to select a cell from which the user will try to register with the network (this is the procedure *IMSI attach*), in order to be able to initiate or receive network calls. These operations are described in Chapter 7.

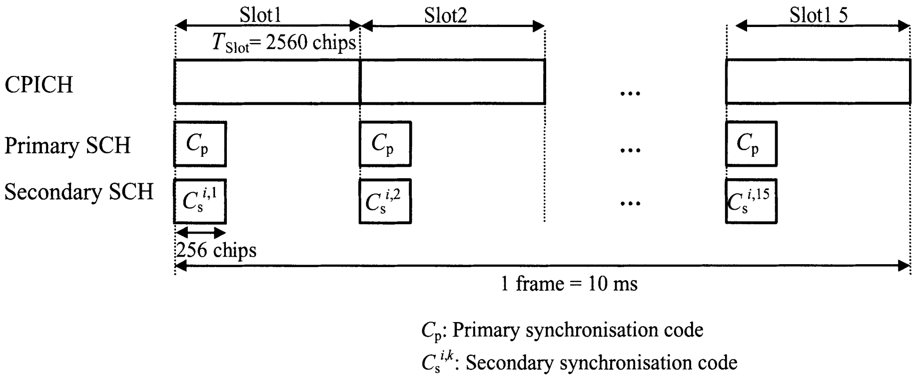


Figure 5.38 Physical channels used in initial synchronisation.

The process described in this section is used by the mobile when it is switched on to search the surrounding UTRAN/FDD cells in the absence of any previous knowledge of the radio environment. This search is made up of three main phases (Figure 5.38) and uses the properties of three physical channels emitted by the base stations in all UTRAN cells:

- the primary synchronisation channel (P-SCH);
- the secondary synchronisation channel (S-SCH);
- the common pilot channel (CPICH).

First Phase: Slot Synchronisation

The primary synchronisation channel is made up of a sequence of 256 chips, repeated on every occurrence of the slot and transmitted to all the cells in the UTRAN network. This sequence, specified in the UTRAN standard, is characteristic of all cells using the UTRAN-FDD technology. It is the same for all UTRAN cells, whether or not they belong to the same network operator.

When it is switched on, the mobile tries to get slot synchronisation with the nearest base station, i.e. the one whose P-SCH is received most strongly. For that, the mobile uses the properties of the P-SCH, by correlation with a copy of the code C_p used by the network, scanning the whole FDD frequency band.

Second Phase: Frame Synchronisation

The secondary synchronisation channel uses a sequence made up of 15 secondary codes of 256 chips broadcast by the network in step with the primary synchronisation channel: (C_s^{i1}, C_s^{i2}, . . . C_s^{i15}).

The UTRAN standard defines a total of 64 sequences C_s^i, constructed in such a way as to ensure that no sequence becomes equal to another by simply shifting the 256 chips used in each sequence. Each sequence is thus strictly unique and allows a cell to be identified uniquely within a given geographic zone. Once the slot synchronisation has been set up, the mobile goes to find the sequence C_s^i used by the cell during the first phase. When this sequence has been recovered, the mobile knows the synchronisation frame of the cell, and hence knows when the frames on the common control channels sent to the cell and synchronised with the P-SCH are sent.

Since a cellular communication network usually contains more than 64 cells, the sequences have to be reused. When configuring the network, the operator must therefore be careful to keep an adequate distance between cells using the same sequence, in order to avoid any ambiguity for the mobile.

Third Phase: Finding the Primary Scrambling Code

The common pilot channel (CPICH) is a common physical channel broadcast to each cell by the network and made up of a sequence of symbols known by the network and the mobile. These symbols are scrambled with the help of the primary scrambling code used by the cell. Finding the primary scrambling code is an important operation for the mobile because this code is also used by the P-CCPCH for broadcasting system information to the cell.

There are 512 primary scrambling codes, divided into 64 groups of eight. The UTRAN standard defines a one-to-one correspondence between the 64 groups of eight scrambling codes and the 64 C_s^i possible secondary synchronisation channel sequences. Once the synchronisation frame has been acquired, therefore, the mobile can determine the group of codes to which the primary scrambling code used by the cell belongs. To recover the code actually used, the mobile carries out a correlation between the data received by the CPICH of the cell and each of the eight possible codes.

Remarks

Because of the number of carrier frequencies contained in the UTRAN-FDD band and the time needed for calculation and for acquiring the different channels involved, the time necessary for the initial cell search is far from negligible. We must realise, however, that these operations are only really necessary when the mobile is powered up. One it has selected the initial cell, information about neighbouring cells will be available to it over the P-CCPCH channel. Hence the cell reselection phases, made necessary by the physical movement of the mobile in the network, will be much simpler. Furthermore, the operations described above represent only the very first step in selecting the initial cell. This step must be complemented by reading the system data broadcast by the cell, which allows the mobile to know whether it can really consider the cell to be suitable from the point of view of the user; in particular, if the cell belongs to a network that is not accessible to the user, it must be discarded.

5.3.2 Broadcasting System Information

The network uses the primary common control physical channel (P-CCPCH) to broadcast information about the network configuration periodically to the cell. The

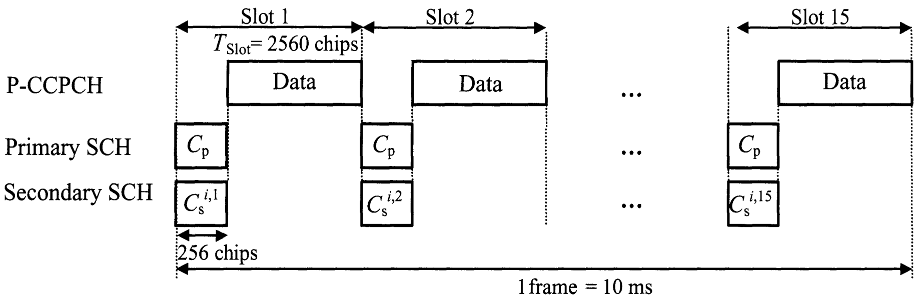


Figure 5.39 Structure of the P-CCPCH.

P-CCPCH is the equivalent of the beacon channel (BCCH) in GSM. The system information constitutes an important body of data, including inter alia:

- network information (the identity of the PLMN, etc.);
- data about the current cell (maximum authorised transmission power, structure of the common channels, etc.);
- information about adjacent cells (carrier frequency, primary code, etc.);
- information on neighbouring technologies (list of adjacent cells of GSM or cdma2000 type, etc.).

Figure 5.39 shows the structure of the P-CCPCH. It used the OVSF code $C_{256,1}$. Since it is not transmitted during the first 256 chips of each slot, it can only carry 18 bits of information per slot, all of which can be used by data coming from the transport channel BCH.

A quick calculation shows that the useful data rate of a P-CCPCH is around 13.5 Kbps:

$$18 \text{ bits} \times 15 \text{ slots}/10 \text{ ms} = 27 \text{ Kbps/s}$$

Used with a convolutional code of code rate ½, the useful throughput of the P-CCPCH is thus approximately 27/2 = 13.5 Kbps. By way of comparison, the throughput of the GSM equivalent of the P-CCPCH is around 780 bps (23 useful bytes per 235 ms multi-frame).

The low capacity of the BCCH is a significant limitation of the GSM standard. The new services introduced into the GSM standard, such as GPRS, HSCSD and EDGE, make important demands on BCCH broadcasting, which can only be satisfied at the expense of the effectiveness of certain mechanisms, and thus at the expense of the quality of service provided to the user.

The extra capacity provided by the P-CCPCH is not otiose. It guarantees a certain longevity to the mechanisms currently defined in the UTRAN standard when new services are created.

5.3.3 Paging

The secondary common control physical channel (S-CCPCH; Figure 5.40) is used by the UTRAN to support at the same time the forward access channels (FACH) and the paging channel (PCH).

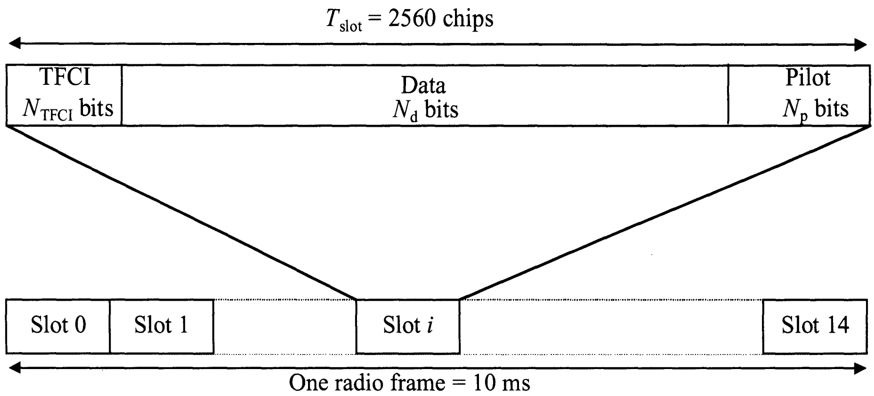


Figure 5.40 Structure of the S-CCPCH.

The S-CCPCH contains two types of information:

- data bits, i.e. either data packets or signalling messages destined for a user (in the case of the FACH) or paging messages (in the case of the PCH);
- physical layer control information: pilot bits and TFCI bits indicating the format of the data part of the S-CCPCH frame.

The S-CCPCH can take any of 18 different formats, depending on the spreading factor chosen by the network and the number of pilot and TFCI bits used. Depending on the values of these different parameters and the type of channel coding used, the useful data rate of the S-CCPCH can very from ten to several hundred kilobits per second.

5.3.4 Call Set-up

When the mobile wants to set up a connection with the network, in response to a paging message or at the request of the user of the mobile, the mobile must make the request on a special physical channel, the physical random access channel (PRACH).

The initial message sent by the mobile is made up of two parts:

- a header or preamble of 4096 chips;
- the message part proper, from 10 to 20 ms, which contains data bits (the DPDCH) and level one (the DPCCH) signalling bits.

In the case of a network access, the DPDCH contains the message RRC CONNECTION REQUEST (see Chapter 7).

The PRACH is a collision channel, in the sense that several mobiles may possibly send a RACH header at the same moment. In order to minimise the probability of a collision between two headers, PRACH users are distributed by time (there are 15 access slots or header transmission windows every 20 ms) and by code (the mobile randomly chooses a signature – a 16-bit code – from among 16 possibilities, this signature being used to generate the header).

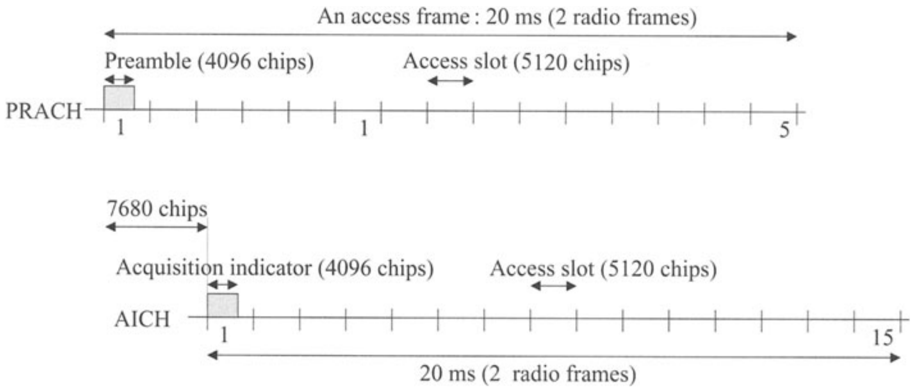


Figure 5.41 RACH and AICH access slots.

The PRACH access slot is twice as large (5120 chips) as a normal slot. The 15 access slots of the PRACH are thus distributed over 20 ms, i.e. two radio frames (Figure 5.41).

Once the header has been received correctly by the network, an acquisition indication is sent by the Node B on a special physical channel called the acquisition indicator channel (AICH). Like the PRACH, the AICH is divided into 15 slots of 5120 chips. In order that each mobile can identify the acquisition indicator meant for it, this indicator is generated using a signature that depends on that used by the mobile for generating the header.

Figure 5.42 shows an example of the transmission of a message on the PRACH. Since the mobile does not know how far it is from the antenna of the base station, the header is initially sent with the minimum power level, and is then repeated with steadily increasing power levels until an acknowledgement is received from the network. One the acknowledgement has finally been received, the mobile can send the message.

The gap between two repetitions of the header, a minimum of three access slots, is chosen randomly in order to minimise the chances of collision during the repeat of the header.

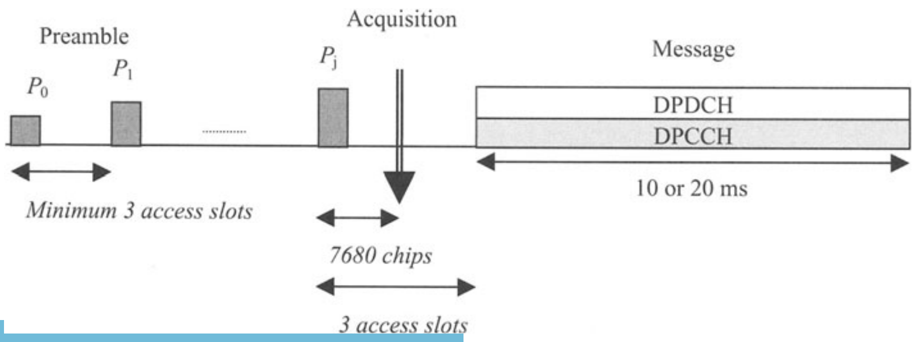


Figure 5.42 Example of a message being sent on the BRACH.

Table 5.4 Formats used by the DPDCH part of the PRACH.

Format number	Bit rate (kbits/s)	SF	Bits per frame	Bits per access slot
0	15	256	150	10
1	30	128	300	20
2	60	64	600	40
3	120	32	1200	80

Table 5.4 shows the different formats that the DPDCH part of the PRACH can take. Depending on the spreading factor (SF) used in the cell on the PRACH, a 10 ms message can contain from 150 bits (format 0) to 1200 bits (format 3), i.e. from 9 to 75 useful bytes (the PRACH uses a channel coding with a code rate of $\frac{1}{2}$).

We will see, in the chapter on call set-up, that the access capacity of the PRACH can be split into sub-parts depending on the class of service allocated to the user. This feature of UMTS allows the probability of collision to be reduced for certain services, such as emergency calls.

5.3.5 Data Transport on a Dedicated Transport Channel

The DPCH is one of the basic physical channels used by the UTRAN for the transport of user data. The DPCH is a resource dedicated to a single user, offering a constant data rate and propagation delay over the radio interface.

The pilot bits are a set of bits known to the sender and receiver, intended to help reception (the pilot bits are used for channel estimation to facilitate coherent detection). They are multiplexed both in code and in time; in other words, they are included in each slot.

The DPCH is, in fact, made up of two physical sub-channels, the dedicated physical data channel (DPDCH) and the dedicated physical control channel (DPCCH). In the context of the physical layer, the concepts of *control channel* and *data channel* do not have quite the same meaning as in radio protocol layers, where we have met them before.

The DPDCH is used to carry all the data coming from higher layers of the protocol. We therefore find in the DPDCH user level data (voice, messages, etc.) as well as signalling data associated with the control level, such as call signalling or RRC signalling.

The control data carried by the DPCCH is specific to the physical layer of the radio interface and is not therefore transferred to the radio protocol layers. The following data items are carried by the channel:

- the pilot bits;
- the transmission power control (TPC) bits;
- the bits that make up the transport format combination indicator, indicating the transport format used by the transport block or blocks contained in the frame;
- the feedback information bits, used only in the uplink direction.

Figure 5.43 shows the format of a downlink DPCH.

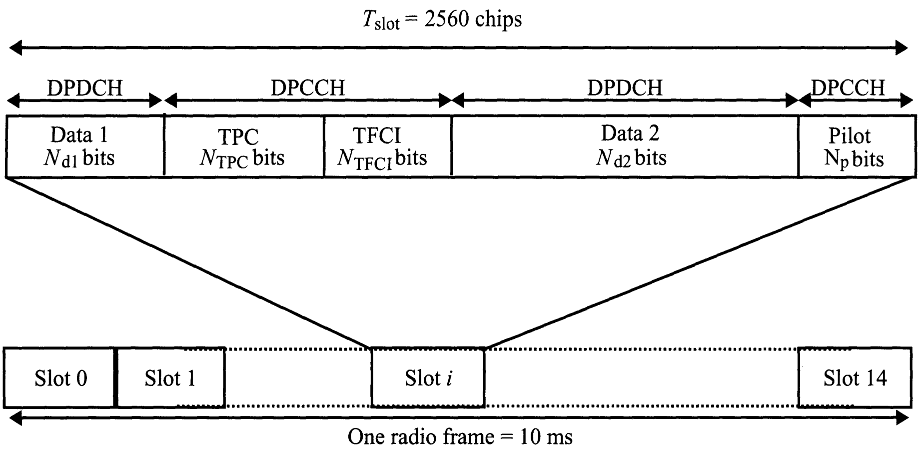


Figure 5.43 Structure of the downlink DPCH.

The standard defines 49 different formats for the downlink DPCH frame, depending on the number of bits reserved for the transport channels and on the spreading factor used.

Table 5.5 shows the parameters of a few of these formats.

Format 1 corresponds to the lowest data rate (two data bits per slot, hence 30 bits per 10 ms frame). The useful data rate of this channel is less than 1.5 Kbps if channel coding with a code rate of ½ is used.

At the other end of the scale, format 16 allows the highest user data rate (1248 data bits per slot) to be attained. This channel can provide a useful data rate of the order of 936 Kbps with a channel code rate of ½. When a channel with this performance is used in a cell, the capacity of the cell is very much reduced, because of the transmission power necessary to transmit the data on the channel and the fact that the code used, with a spreading factor of 4, is situated high in the tree.

Figure 5.44 shows the format of the DPCH in the uplink direction.

In contrast to the downlink DPCH, the DPCCH and DPDCH sub-channels are separated on to the I and Q paths of the transmission chain. To each data bit sent

Table 5.5 Examples of the downlink DPCH formats.

Format number	Bit rate (kbits/s)	Symbol rate (ks/s)	SF	Bits/ slot	DPDCH (bits/slot)		DPCCH (bits/slot)		
					N_{D1}	N_{D2}	N_{TPC}	N_{TFCI}	N_{Pilot}
0	15	7.5	512	10	0	4	2	0	4
1	15	7.5	512	10	0	2	2	2	4
4	30	15	256	20	2	12	2	0	4
5	30	15	256	20	2	10	2	2	4
10	60	30	128	40	6	24	2	0	8
15	960	480	8	640	120	488	8	8	16
16	1920	960	4	1280	248	1000	8	8	16

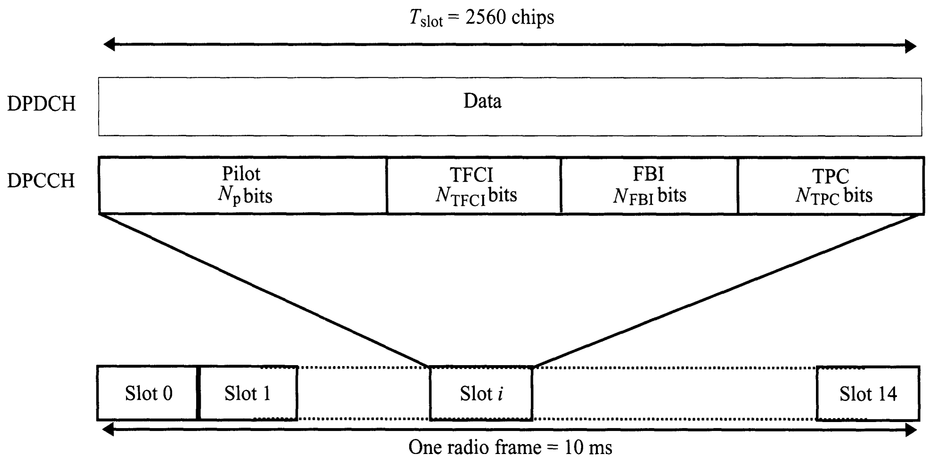


Figure 5.44 Structure of the uplink DPCH.

on the I path, there corresponds a symbol, i.e. the bit rate and the symbol rate for the DPDCH are the same.

Table 5.6 gives the parameters for each of the six uplink DPDCH formats. The maximum data rate for an uplink DPDCH is 480 Kbps of useful data, using format 6 with a channel coding rate of ½.

5.3.6 Data Transport on the FACH

In Section 4.4.1, we discussed the states of the RRC CELL_FACH connection, using which the user data is transmitted from the network to the mobile over the common transport channel FACH (Forward Access Channel). This is the transmission mode that most closely resembles data transmission in the second generation GPRS networks, to the extent that no prior resource reservation is carried out. The channel is used on demand, depending on the instantaneous data transmission requirements.

This section describes how data is transmitted on the FACH and, in particular, the multiplexing technique used.

The physical channel supporting the FACH is the secondary CCPCH (S-CCPCH). It is the same type of physical channel as is used for transmitting paging messages,

Table 5.6 Formats used by the uplink DPDCH.

Format number	Bit rate (kbits/s)	Symbol rate (ks/s)	SF	Bits per frame	Bits per slot	N_d
0	15	15	256	150	10	10
1	30	30	128	300	20	20
2	60	60	64	600	40	40
3	120	120	32	1200	80	80
4	240	240	16	2400	160	160
5	480	480	8	4800	320	320
5	960	960	4	9600	640	640

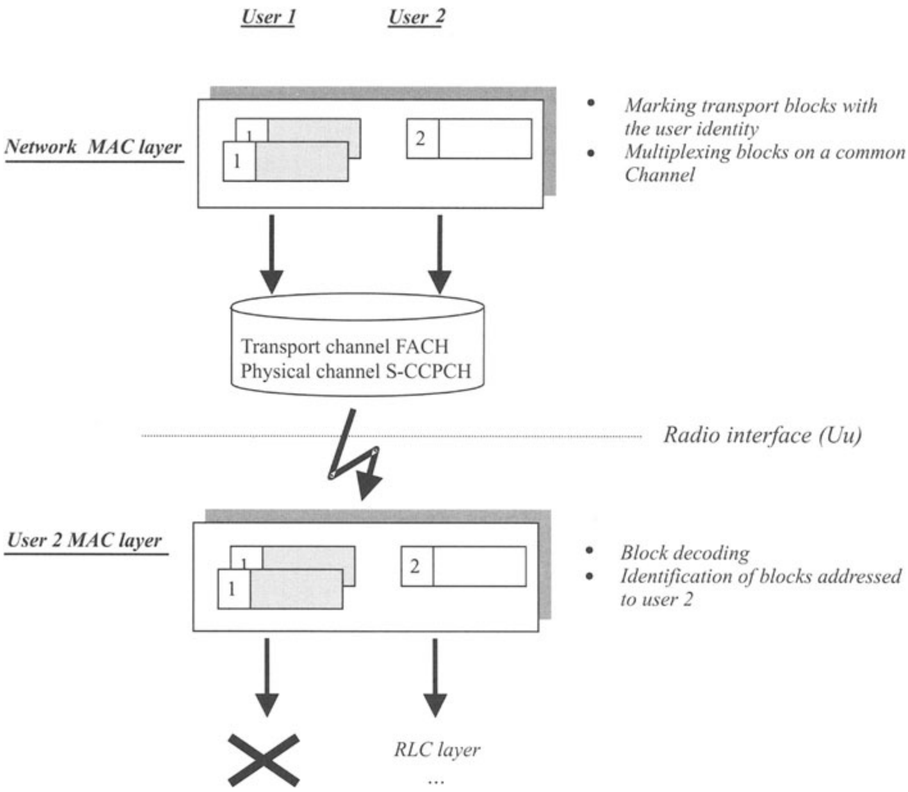


Figure 5.45 Data transmission on the FACH transport channel.

described in the previous section. Figure 5.45 shows how data is transmitted on the FACH.

The MAC layer of the network collects the data blocks intended for transmission over the FACH. Each packet is then marked with a special header containing the identity of the mobile to which it is addressed, before being multiplexed and transmitted over the physical channel.

At the level of the mobile, each frame that the network puts on to the S-CCPCH is decoded by the mobile, in order to re-assemble the MAC blocks sent by the network. The MAC layer of user 2's mobile then sorts the blocks received, in order to transmit to the higher layers (RLC, PDCP, RRC, etc.) only the blocks actually intended for user 2.

5.3.7 Data Transport on the DSCH

Just like the FACH, the downlink shared channel (DSCH) allows different transmissions from the network to the mobile to be multiplexed over a shared resource. The DSCH, however, offers certain advantages in comparison to the FACH:

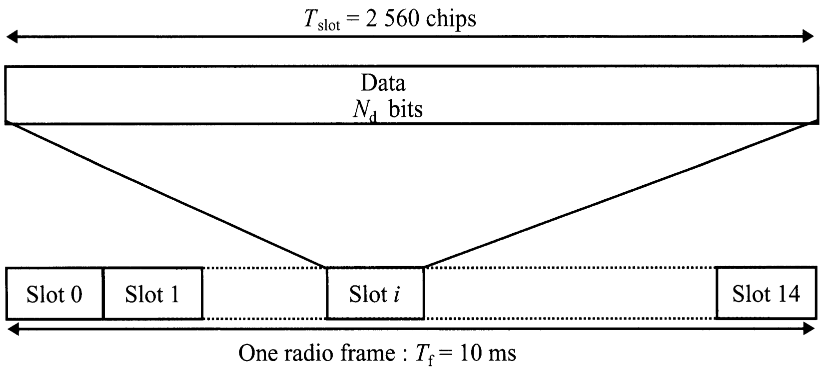


Figure 5.46 Structure of the PDSCH.

- The presence of a control channel associated with the DSCH allows the mobile to avoid systematically decoding the blocks transmitted on the common channel and thus to reduce battery use.
- The DSCH can be considered to be a variable size FACH, i.e. no resources are tied up when it is not being used.

In contrast to the FACH, the DSCH can only be active if the mobile is in the RRC CELL_DCH state. In this state, uplink data is transmitted over a dedicated transport channel (DCH).

Structure of the DSCH

The DSCH transport channel uses the physical downlink shared channel (PDSCH), whose structure is described in Figure 5.46.

Table 5.7 shows some of the different possible PDSCH formats defined in the standard. Depending on the format used, the PDSCH allows user data rates of between 15 Kbps and 960 Kbps approximately, assuming a channel code rate of ½.

This channel transmits only user data. Physical layer control data (including transport format combination indicator bits and transmission power control bits) is transmitted across the control channel associated with the DSCH.

As shown in Figure 5.47, the principle used in the associated control channel is to cut the TFCI field carried by the DPCCH into two parts:

- the first part, called TFCI (1), indicates to the mobile the coding format of the DPDCH part of the associated channel;

Table 5.7 Examples of the PDSCH formats.

Format number	Bit rate (kbits/s)	Symbol rate (ks/s)	SF	Bits per frame	Bits per slot	N_d
0	30	15	256	300	20	20
3	240	120	32	2400	160	160
6	1920	960	4	19200	1280	1280

- the second part, called TFCI (2), specifies the user data format transmitted over the PDSCH.

Figure 5.47 shows an example of the use of the DSCH for a network user. In the example, two logical channels are used:

- a DCCH used to transmit signalling information connected with the call. The information transmitted by this channel will be supported by the associated dedicated transport channel;
- a DTCH used to transmit the user data, which will be transmitted over the common transport channel DSCH.

The MAC layer of the network switches the PDUs coming from each logical channel to the selected transport channel.

On receiving the data transmitted across the associated DCH, the mobile must decode it in order to decode correctly the data transmitted over the DSCH.

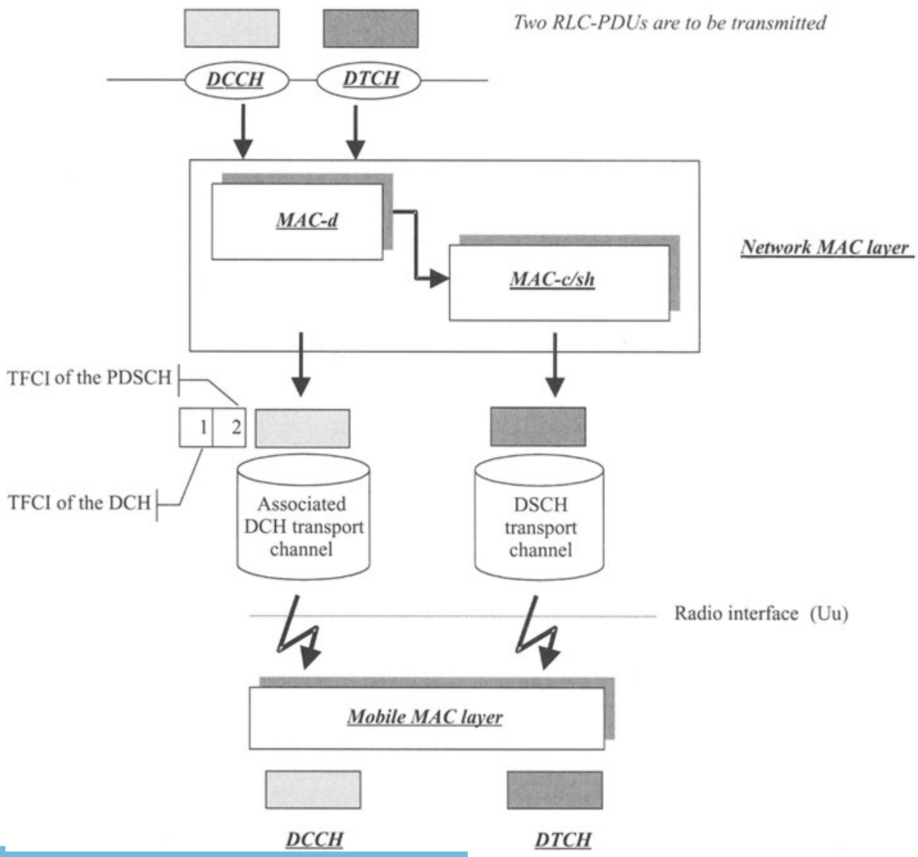


Figure 5.47 Data transmission on the DSCH transport channel.

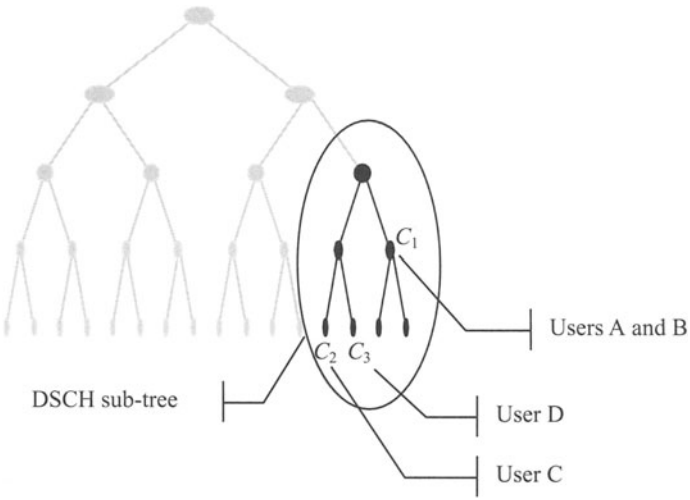


Figure 5.48 Example of the allocation of DSCH codes.

Multiplexing on the DSCH

In the same way as the FACH, the DSCH is a common transport channel in the sense that the resources provided by the channel can be used by several users. The DSCH has the special feature that user data can be multiplexed in two ways: in code or in time.

As shown in Figure 5.48, the DSCH can be considered as a sub-tree of the code tree. The network thus has the choice of using the code at the head of the sub-tree for all the users or of using one of the codes of the sub-tree for a subset of the users or of using simultaneously for n users the n codes situated in the branches of the tree.

In the example, suppose that code C_1 is used by users A and B of the DSCH and that codes C_2 and C_3 are reserved respectively for users C and D. Under this hypothesis, the DSCH transmission for users A, B, C and D might resemble the schema shown in Figure 5.49.

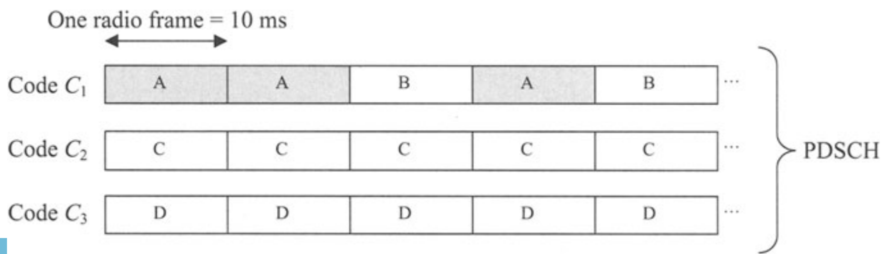


Figure 5.49 Example of multiplexing users.

Users A and B have to share the resource carried by the code C_1. The data for these two users is therefore time division multiplexed over the DSCH. In contrast, since each of users C and D have their own code from the tree, their data can be transmitted simultaneously.

The advantage of the DSCH in comparison with a dedicated channel lies in the fact that, when the DSCH user has no more data to transmit, it is possible to redistribute the resource to another user very rapidly, in time for the next frame, even.

5.4 Some Examples

By way of illustration and synthesis of what we have said about the radio interface in Chapters 4 and 5, in this section we will present two examples of multiplexing and coding corresponding to applications which will be among the first to become available when the commercial exploitation of UMTS begins:

- data transmission in circuit mode at 64 Kbps, used for example to support videophones;
- conventional voice telephony in the AMR12-2K mode.

As stated in Chapter 3, in the discussion of the basic concepts of UMTS, the UTRAN standard does not in any way specify the precise parameters of resources used in providing the different user services. These examples are, however, fairly representative of reality because they are taken from the document 3GPP 34.108, which species the conformance tests applicable to UMTS mobiles.

5.4.1 The Services and the OVSF Tree

We have seen in the previous sections that the network allocates an OVSF code to each service requested, the position of the code in the tree depending on the data rate and the quality of service required. Figure 5.50 shows several examples of the use of codes from the tree for downlink communication.

The lowest level codes (SF = 512) offer a very low useful data rate. They are generally reserved for special uses, such as short messages, sign-on procedures, or updating of positioning information. These procedures are addressed in detail in Chapters 7 and 8.

The Amr mode, which will be used for voice telephony applications, supports codes with SF of 64 or 128. The second alternative allows the capacity of the network to be increased, at some cost to the speech quality.

As the processing gain diminishes, it is possible to increase the useful data rate, up to a maximum of 2 Mbps, the use of an SF less than 4 not being possible.

Although envisaged by the UTRAN standard, the transmission of user data at 2 Mbps remains only a theoretical possibility. It would require, in fact, three physical channels with an SF of 4, used simultaneously. The last code with SF = 4 being itself partly used by the common channels of the cell (CPICH, P-CCPCH, etc.), it is in practice impossible to allow an additional user into the cell.

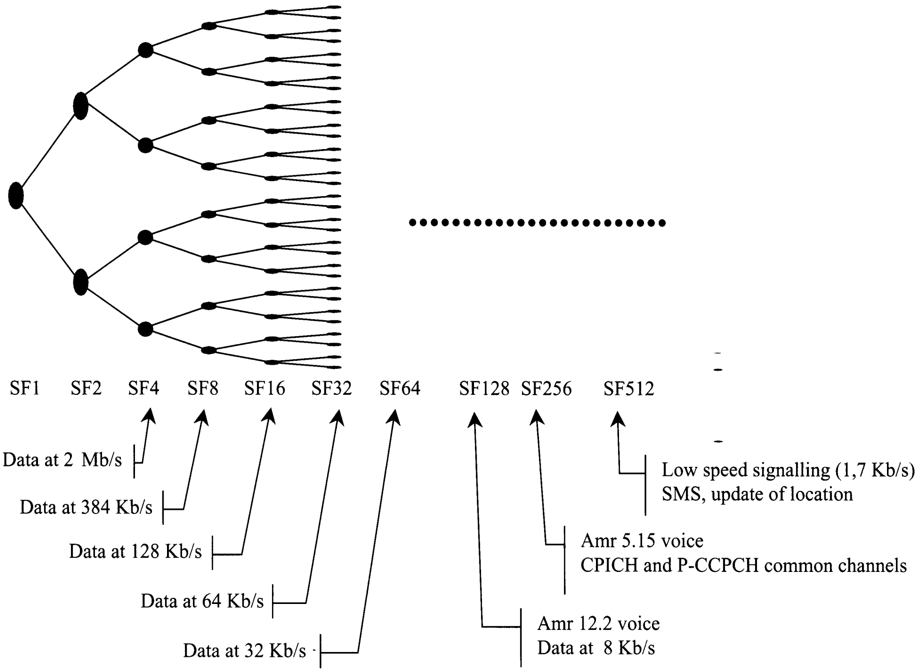


Figure 5.50 Processing gain and corresponding services.

5.4.2 The 64 Kbps Data Circuit

The example described in this section shows the operations carried out by the radio layers of the UTRAN (level 1 and layers of the level 2 protocol) in the case of a circuit-switched data transmission service operating at 64 Kbps.

Figure 5.51 show the different channels that will be employed in the call.

The user data (for example, voice or image data) is transmitted over the radio interface across a single dedicated logical channel: the dedicated traffic channel

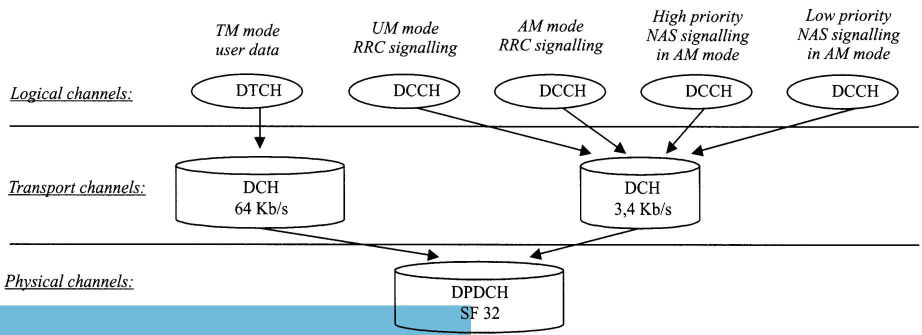


Figure 5.51 Radio channel multiplexing.

(DTCH). In the case of an application with strict real-time constraints such as videophone, the radio interface will not re-transmit lost or erroneous data. The RLC protocol layer will therefore be configured in transparent mode.

Unlike the flow of user data, there may exist different types of signalling messages that will be exchanged during the course of the call. In particular, the UTRAN standard distinguishes:

- RRC signalling transmitted in both unacknowledged mode and acknowledged mode;
- high priority NAS signalling, corresponding to exchanges between the MM and CC protocol layers;
- low priority NAS signalling, including for example short messages.

These different signalling flows will be multiplexed over the same dedicated transport channel by the MAC.

The signalling information and the user data will be carried by different transport channels, illustrating the fact that that the quality of service required for each is different. (In particular, we will see that the signalling information is better protected than the user data.) Table 5.8 gives the principal features of the two transport channels used and Figure 5.52 shows the operations that will be carried out by the physical layer for each of the transport channels.

The error control operations (adding the CRC and channel coding), data rate matching and interleaving are all applied separately to each channel. The CRC bits are added to each transport block before the channel coding (turbocode or convolutional) is applied. The data rate matching is then carried out, depending on the maximum number of useful bits that can be carried by the 10 ms radio frame. The blocks resulting from this operation are then interleaved and segmented according to the value of the TTI of each transport channel.

Since the TTI of the speech transfer channels is 20 ms, two radio frames will be necessary to carry one speech frame produced by the transcoder. The transport blocks of the speech channels will therefore be cut into two segments. The TTI of the signalling channel is 40 ms in our example. A DCCH transport block will therefore be cut into four segments, each one carried by a 10 ms radio frame.

The two transport channels are then multiplexed on to one frame of a single physical channel, segmentation into physical channels not being used in this case.

As shown in Table 5.9, the physical channel uses a code with spreading factor 32, providing 140 bits per slot. (A slot contains 2560 chips, that is 2560/32 = 80 QPSK

Table 5.8 Characteristics of the transport channels.

	DTCH	4 × DCCH
Data rate	64 Kbps	3.4 Kbps
Transport block size	2 × 640 bits	148 bits
TTI	20 ms	40 ms
CRC	16 bits	16 bits
Channel coding	turbocode	convolutional ⅓

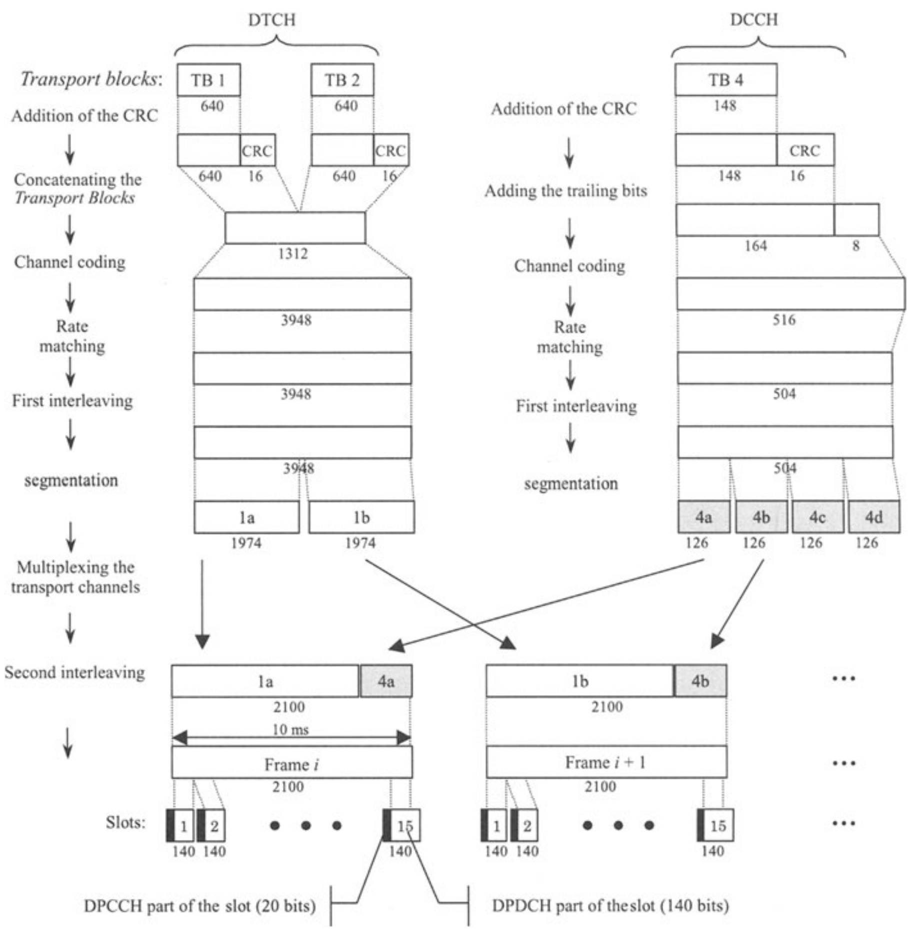


Figure 5.52 Physical layer operations.

symbols = 160 bits of information.) Of the 160 bits, 20 are reserved for the DPCCH, i.e. used to transport physical layer control information:

- 8 bits for the pilot signal;
- 4 bits for transmission power control;
- 8 bits for coding the TFCI.

Table 5.9 Configuration of the physical channel.

Data rate (Kbps)	SF	Bits/slot	DPDCH bits	DPCCH bits	Number of slots per frame
240	32	160	140	20	15

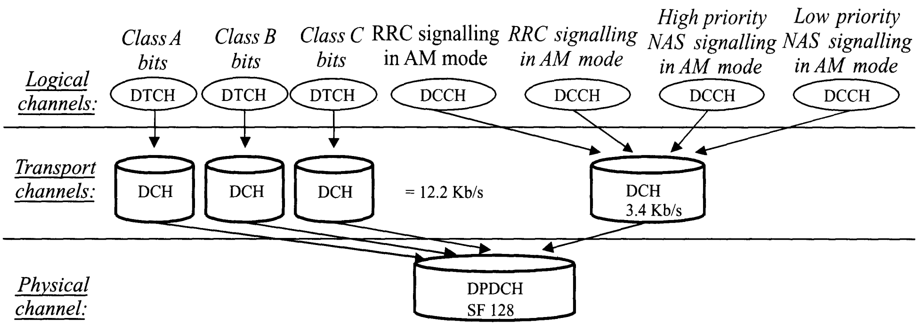


Figure 5.53 Multiplexing the radio channels.

5.4.3 Telephony

The example given in this section is that of a voice service using the AMR12-2K mode. This mode, already described in Chapter 4, is one of the modes most widely used for speech transmission.

In the example in Chapter 4, we saw that the speech frame coming out of the transcoder was made up of three classes of bits: A, B and C. In the example below, these three classes of bits are considered to form part of three distinct transport channels, allowing the physical layer to apply specific treatment to each of the classes.

Figure 5.53 demonstrates the multiplexing of the different channels used in this example and Table 5.10 shows the principal characteristics of the four transport channels, indicating the specific channel coding associated with each class of bit. Figure 5.54 shows an example of the operations carried out by the physical layer in the case of the four transport channels of Table 5.10.

As in the previous example, the transport blocks for the channels using a TTI of 20 ms are cut into two segments in order to be carried on two successive radio frames. The DCCH processing is identical to the example in the previous section.

As Table 5.11 shows, the physical channel uses a code with spreading factor 128, providing 40 bits of information per slot. (A slot contains 2560 chips, that is 2560/128 = 20 QPSK symbols = 40 bits of information.) Of the 40 bits, 6 are reserved for the DPCCH, i.e. used to transport physical layer control information:

- 8 bits for the pilot signal;
- 2 bits for transmission power control.

Table 5.10 Example of the configuration of transport channels.

	DTCH (A)	DTCH (B)	DTCH (C)	4 × DCCH
Data rate		12.2 Kbps		3.4 Kbps
Transport block size	81	103	60	148
TTI	20	20	20	40
CRC	12	0	0	16
Channel coding	convolutional 1/3	convolutional 1/3	convolutional 1/2	convolutional 1/3

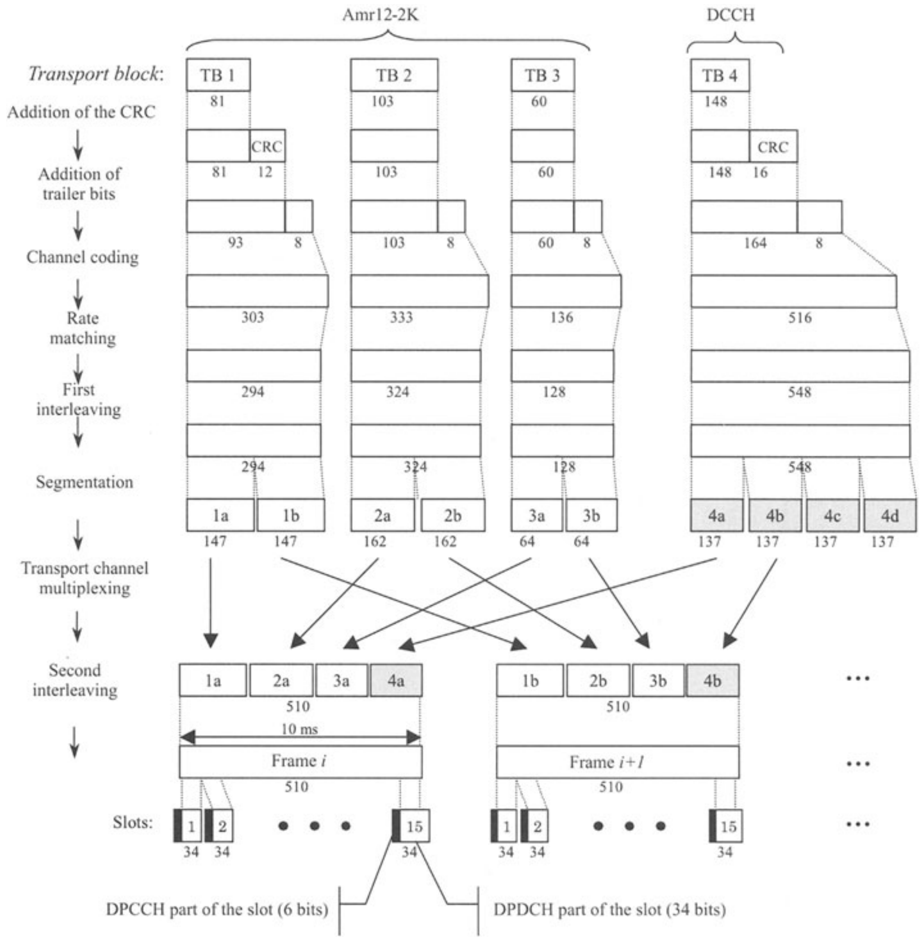


Figure 5.54 Coding and multiplexing.

Table 5.11 Example of physical channel configuration.

Data rate (Kbps)	SF	Bits/slot	DPDCH bits	DPCCH bits	Number of slots per frame
60	128	40	34	6	15

5.5 Timing Relationship Between the Physical Channels

Figure 5.55 shows the temporal relationship between the different physical channels transmitted by the Node B.

With the exception of the S-CCPCH (used by transport channels PCH and FACH) all the common channels of the cell are aligned over the 10 ms frame structure:

- the pilot channel;
- the synchronisation channels (SCH);

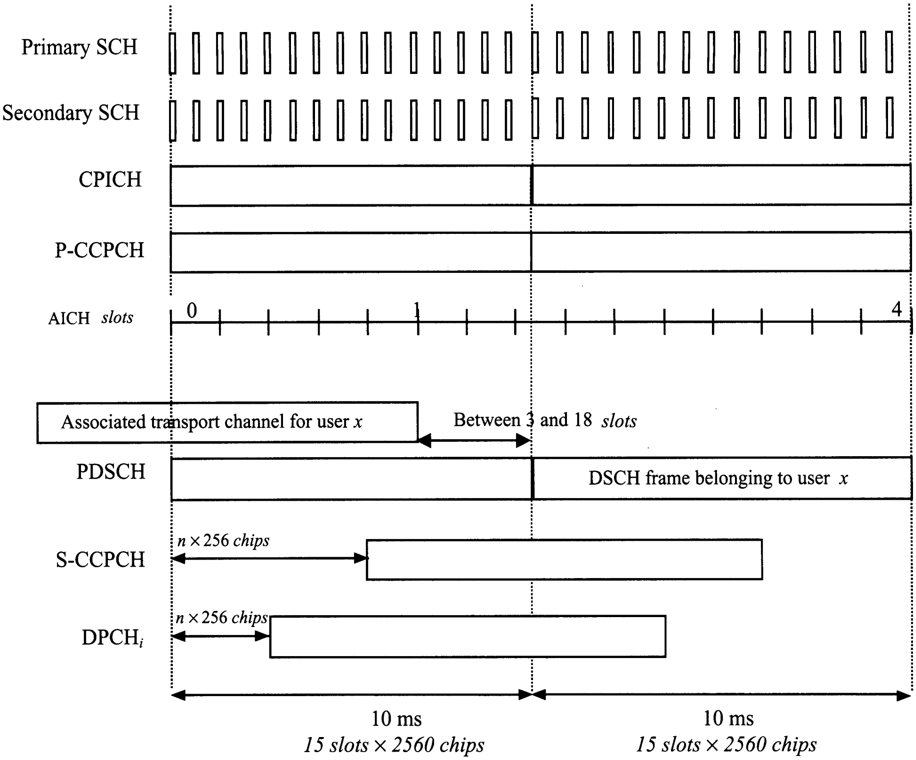


Figure 5.55 Temporal relationship between the Node B physical channels.

- the AICH;
- the PDSCH used by the common transport channel DSCH.

The dedicated physical channels (DPCH) are shifted in time by a multiple of 256 chips, 256 chips being a tenth of a slot.

Figure 5.55 also shows the temporal relationship between the dedicated transport channel associated with the DSCH of a given user, and the data frame transmitted over the PDSCH for that user.

6 The UTRAN Network Protocols

6.1 Introduction

In the previous chapters we have studied the make-up of the UTRAN radio interface and its associated protocols. However, the UTRAN is not limited to the radio interface. Because of the large number of network nodes defined in the UTRAN (Node B and RNC), a set of protocols and interfaces belonging to the terrestrial part of the UTRAN has been defined by the 3GPP to allow these different nodes to exchange signalling information and user data.

Figure 6.1 shows the layered structure of the UTRAN network interfaces: the Iu, the Iub and the Iur. This model is generic and applies to all these interfaces.

This complex model has two purposes:

- To keep the data transmitted independent of the technology used for the transmission. This independence will make it possible for future versions of the UTRAN specifications to use new transmission technologies without affecting the higher layers.
- To keep the user plane and the control plane completely separate. The quality of service required from the transport channels by the user layers is not necessarily the same for user data and control messages. The UTRAN model allows different transport technologies to be defined for the two planes.

For the majority of its interfaces, the GSM standard imposes the transport technology used by the AP layer. Thus, for example, the GSM standard insists that interface A uses 64 Kbps circuits over PCM links. This choice may now appear sub-optimal, but it is in fact consistent with the technology that was in use when this interface was defined.

To change the interface A transport technology now would be a complex undertaking. It would mean extensive changes to the AP protocol (the BSSMAP), whose definition is very dependent on the transport network technology.

In order to avoid such problems, a different approach has been used for the UTRAN network interfaces, ensuring that the UTRAN standard will be more flexible in the face of developments in transport technology. The way in which these interfaces are modelled in fact allows the transport part of the interface to be modified without any impact on the AP layers.

The layered model of the interfaces can be described in terms of either horizontal or vertical sectioning.

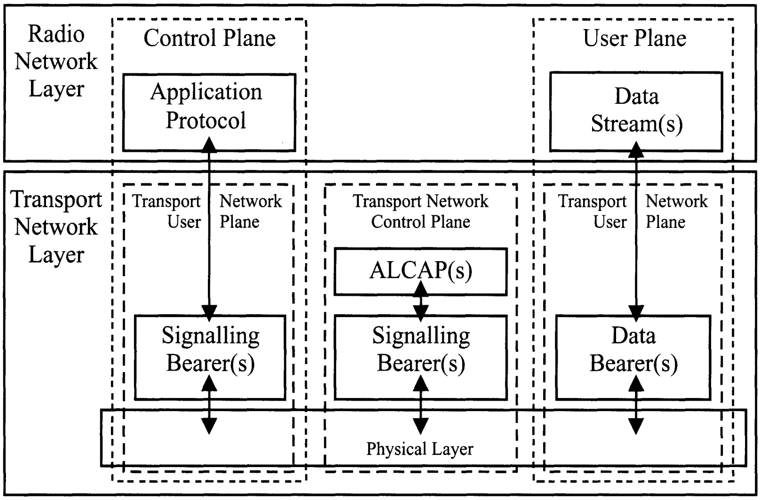


Figure 6.1 Layered model of the UTRAN network interfaces.

Horizontal sectioning leads to two layers:

- The transport network layer. This layer is made up of the physical layer, the communication channels (signalling/data bearer), and the access link control application part (ALCAP), used to set up the transmission routes of the user plane.
- The radio network layer, including the network application protocols and the data stream protocols.

Vertical sectioning leads to three planes:

- The control plane, including the network protocols, i.e. the signalling exchanges between the different pieces of equipment in the UTRAN fixed network.
- The user plane, through which the data exchanged on the interfaces passes.
- The transport network control plane. This plane has the special property of only being present in the transport layer, because it is used to set up the user plane transmission routes (the data bearers). This plane is not always present in the UTRAN interfaces. It exists in the cases in which the user plane channels are pre-established.

The term 'user plane' used in describing the vertical sectioning is a little deceptive, because this plane can also be used to carry control information intended for the user. (This is the case with RRC protocol messages.) In fact, from a transport network point of view, the user plane should be considered as bringing together all the data – in the broad sense – carried by the network interfaces.

The following sections describe the contents of the different boxes in the generic model. The way the different application protocols are used is described in a more concrete manner in Chapter 7.

6.2 Opening up the Interfaces

In defining an interface between two points of the network, two approaches are possible.

The first is to standardise the interface, i.e. to specify through a standardisation group the functions of the interface, the elementary procedures of the protocol, and the exact format of the messages exchanged. The interface is then defined in detail and, as far as possible, unambiguously. It becomes possible to connect equipment from different manufacturers to the interface, thanks to this unique, standard view of the interface. Such an interface is said to be *open*.

In contrast, the second approach is to specify only the functional part of the interface, leaving each manufacturer at liberty to define his own view or implementation of it. We then speak of a *closed* interface. In this case, it is no longer possible to connect equipment from different manufacturers as easily as with an open interface, unless there is some sort of alliance or special agreement between them.

The access network of the GSM standard (the BSS) defines only a minimum number of open interfaces. Only the interfaces between the BSS and the equipment external to the BSS are standardised, i.e. the radio interface (between the network and the mobile) and interface A (the GSM equivalent of the Iu). The other interfaces internal to the BSS (the Abis O&M and the Abis RSL between the BSC and BTS equipment) are defined functionally, but the descriptions of the elementary procedures and the messages, where they exist, are not mandatory. Thus each GSM network supplier has developed his own version of these interfaces.

The approach of the UTRAN standardisation committees is significantly different. All the UTRAN interfaces (not only the interfaces visible from outside the UTRAN but also the internal interfaces) are completely specified and standardised. This is the case therefore for the Iu, Iub and Iur.

The consequence of this approach is much greater flexibility for operators in the deployment of their networks, allowing equipment from different manufacturers to be linked throughout the network. It is, of course, always possible for a given manufacturer to implement, for example, his own version of the Iub interface and then to offer a range of products that are only partly compatible with the UMTS standard.

6.3 Network Protocols

A general view of the network protocols is shown in Figure 6.2.

6.3.1 RANAP Network Protocol

The RANAP (*Radio Access Network Application Part*) protocol is used for signalling exchanges between the core network (i.e. the MSC or the SGSN) and the RNC. From a functional point of view, this protocol is quite similar to the GSM interface A protocol, BSSMAP, between the BSC and the MSC.

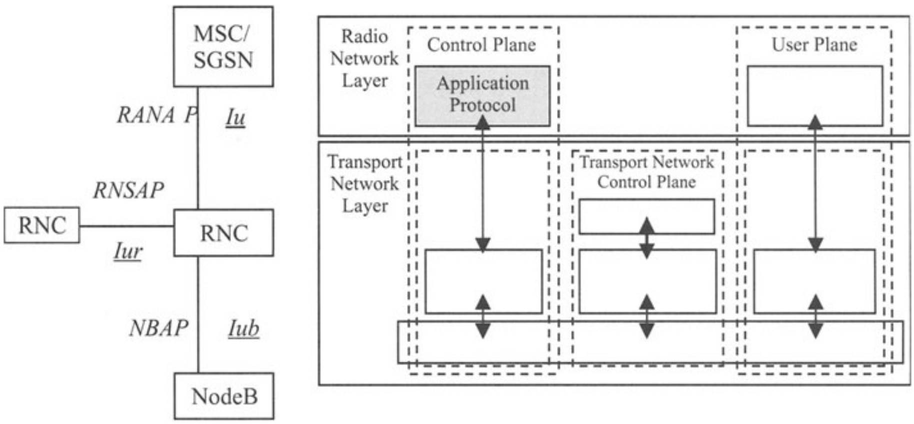


Figure 6.2 General view of the network protocols.

The procedures and messages supported by the RANAP protocol can be divided into two categories: messages addressed to a specific user or connection, and messages addressed to the RNC. The messages addressed to a specific connection serve to implement the following functions:

- management of the radio access bearers, i.e. allocating, de-allocating and reconfiguring them;
- SRNC relocation (i.e. changing to another serving RNC);
- security functions, i.e. the request to change to encrypted mode;
- transmission of signalling information from higher layers. Signalling information coming from the non-access stratum (CC and MM, for example), functionally transparent to the UTRAN, travels over the RANAP protocol.

Messages addressed to the RNC provide the following functions:

- re-initialisation of the RNC or the MSC. This function is used, for example, in the event of a serious software failure;
- paging a mobile (request for a call to a mobile, coming from the network).

6.3.2 NBAP Network Protocol

The NBAP (Node B Application Part) protocol is used for signalling exchanges between the RNC and the Node B. As with the RANAP protocol, there exist two categories of exchanges: those addressed to a specific user, identified at the NBAP protocol level by a CRNC communication context id, and those addressed to the Node B itself.

The first category comprises messages to carry out the following functions:

- management of radio links, i.e. setting up and closing down links, or reconfiguring them;

- management of radio measurements for dedicated channels, i.e. configuration by the RNC of the measurements carried out by the Node B for a specific mobile, and passing these measures back from the Node B to the RNC;
- power control. This function allows the RNC to adjust the power level of the Node B on the downlink channels.

Messages in the second category allow the following functions to be carried out:

- management of the common transport channels from the cell or cells of the Node B, i.e. configuration of the PICH, PRACH, FACH, PCH and RACH;
- configuration of the cell or cells supported by the Node B;
- configuration of the system information broadcast by the cells of the Node B;
- management of radio measurements on the cell's common channels.

6.3.3 RNSAP protocol

The RNSAP (*Radio Network Subsystem Application Part*) is used for signalling exchanges between two RNCs in the UTRAN. As a result, it includes all the functions of the NBAP protocol associated with links in dedicated mode:

- management of radio links;
- management of radio measurements for dedicated channels;
- power control.

Functions relating to the Node B (configuration and management of common channels) are not present in the RNSAP because they are provided by the RNC directly controlling the Node B (the CRNC) by means of the Iub interface.

6.4 Frame Protocols

The application protocols belonging to the user plane are called frame protocols (FP) (Figure 6.3). These protocols are simpler than the network protocols of the control plane because their main function is limited to the transport of data packets from the user plane. We will see, however, that these protocols do also provide other functions.

6.4.1 Frame on the Iu Interface

Transparent Mode

This mode is used, for example, for services set up across the PS domain, which require no special service from the transport protocol. The frame protocol provides no special service, apart from the transport of user data. The frames in transparent mode contain only the user plane data. There is no protocol header and the frames are of variable length.

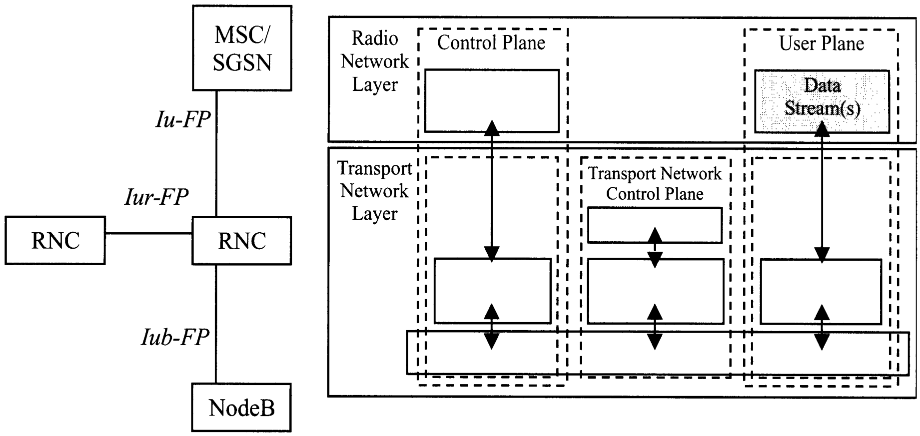


Figure 6.3 General view of the FP protocols.

Non-transparent (Support) Mode

This mode is used when the transfer of user data requires certain special functions from the transport network. In addition to the transfer of user data, the frame protocols of the Iu interface provide the following functions:

- initialisation of the transcoder and control of the data rate. These functions are used in the case of AMR voice communications;
- error handling. This function is used, for example, for detecting erroneous frames.

6.4.2 FP on the Iub and Iur Interfaces

When a drift RNC is present on the communication path between the SRNC and the Node B supporting the radio link or links with the mobile, the drift RNC plays the role of a simple router so far as the user data is concerned.

In addition to the transfer of user data, the frame protocol of the Iub and Iur interfaces provides the following services:

- power control;
- synchronisation of the transport channel.

6.5 Transport Layer

Figure 6.4 shows the transport layer.

6.5.1 Role of the ATM

The ATM (Asynchronous Transfer Mode) is used in the majority of the UTRAN network interfaces. From a very simplified point of view, ATM is a mixed

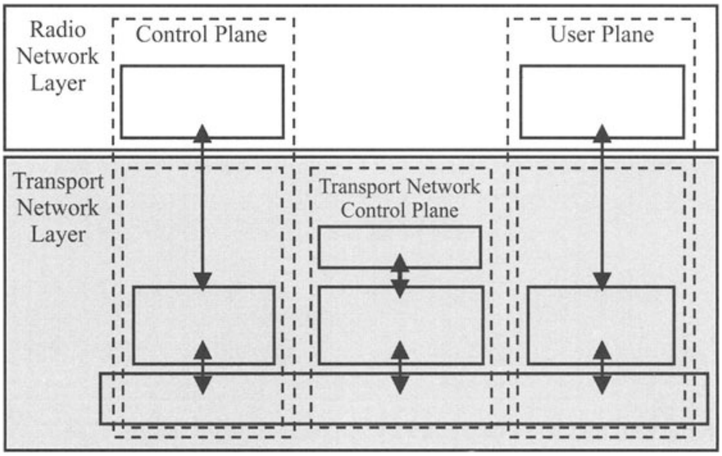


Figure 6.4 The transport layer.

transmission technique that combines the advantages of circuit mode transmission (allowing a data rate and a transmission delay to be guaranteed) with those of packet mode (i.e. the possibility of statistical gains obtained by multiplexing different users with different traffic profiles).

The following sections describe briefly the ATM transmission technique and the use that the UTRAN transport network makes of it.

Asynchronous Transfer Mode

The GSM transport network uses a synchronous transfer mode. The physical transmission medium is divided into frames, themselves cut up into time intervals. One or more time intervals in the frame is allocated to each user of the transport network. The problem with this technique is that the transmission resource is wasted if the user has nothing to transmit.

ATM also uses a transmission medium divided into frames. The difference lies in the fact that the time intervals are common property and are usable on demand by any user, which allows the transmission medium to be used more efficiently. This difference is illustrated in Figure 6.5.

Since the correspondence between the user and the transmission delay has been lost, it is necessary to have recourse to a marking mechanism for the information transmitted. This marking is provided by the header of the ATM cell, the elementary unit of transmission in ATM networks.

Virtual Circuits

ATM operates in connected mode, which means that the transmission path must be set up before the data transfer takes place. The path can be set up dynamically, either by signalling exchanges (in which case the term *switched virtual circuit* is

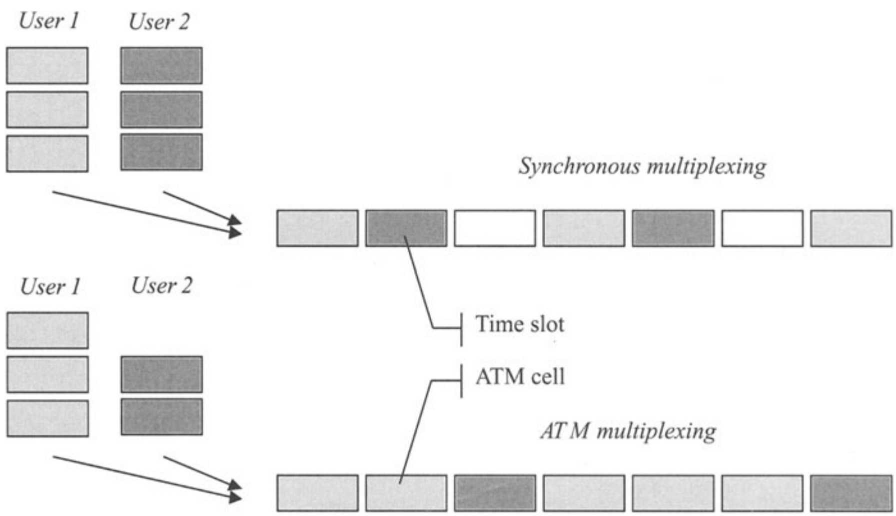


Figure 6.5 Synchronous and asynchronous multiplexing.

used) or by using permanent network connections (*permanent virtual circuit*) (Figure 6.6).

The transmission path is identified by the pair VPI/VCI (Virtual Path Identifier/ Virtual Channel Identifier). The double identification allows two different types of switching to be carried out in the ATM exchanges: virtual path switching and virtual circuit switching (Figure 6.7).

ATM Cells

ATM transmission uses fixed length packets called cells. ATM cells are 53 bytes long, five bytes of which are used as a header. Since the time intervals are common property, each ATM cell includes an addressing mechanism (the VPI/VCI fields in the header), identifying the sender of the cell and allowing the network to route the cell to the right receiver (Figure 6.8).

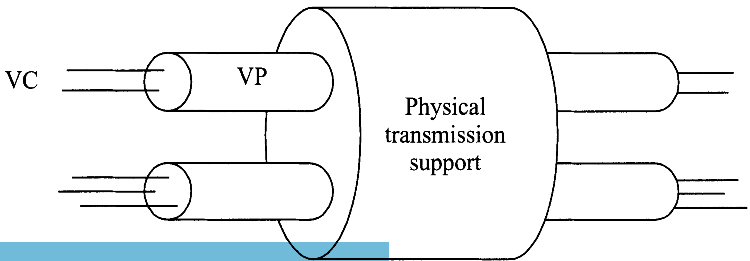


Figure 6.6 Virtual paths and virtual circuits.

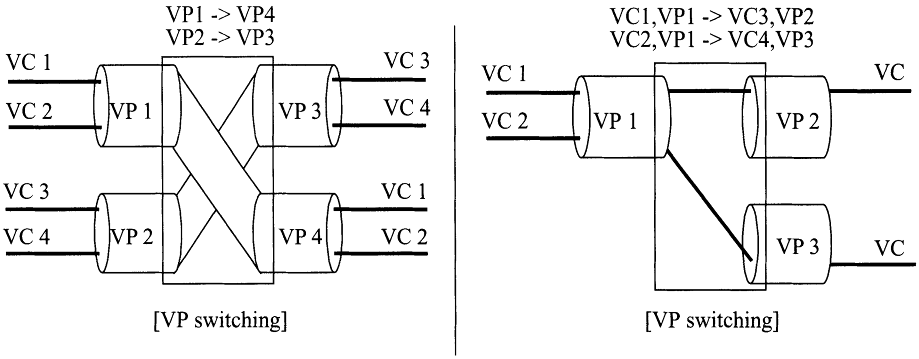


Figure 6.7 Virtual path and virtual circuit switching.

The cell header is made up of the following fields:

- generic flow control (GFC);
- VPI/VCI;
- payload type (PT), which indicates whether the cell contains control information or user data;
- congestion loss priority (CLP), indicating whether the cell can be discarded in the event of network congestion;
- header error control (HEC), which allows error checking to be carried out on the cell header.

ATM Layered Model

Figure 6.9 shows the ATM layered model.

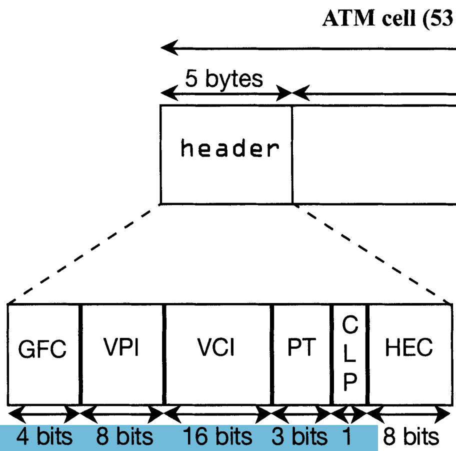


Figure 6.8 Format of an ATM cell.

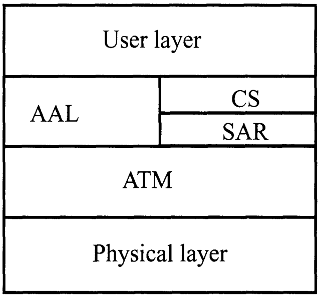


Figure 6.9 ATM layered model.

The principal function of the ATM adaptation layer (AAL) is to provide the quality of service requested by the user. The AAL layer is divided into two sub-layers:

- the convergence sub-layer (CS) provides error detection and correction facilities and carries out end-to-end synchronisation;
- the main function of the segmentation and re-assembly sub-layer (SAR) is to segment the data to be sent into 48-byte packets.

Depending on the quality of service required by the user, ATM offers different types of AAL (AAL1 to AAL5). The two adaptation layers used in the UTRAN transport network are AAL2 and AAL5.

AAL2 is used to carry small packets for applications sensitive to transmission delay. In the case of the UTRAN, user data (data packets, voice, etc.) on the Iub, Iur and (for CS services only) Iu interfaces is carried by AAL2.

AAL5 is used to carry variable length packets for applications that are not very sensitive to transmission delay. In the UTRAN, signalling on the Iu, Iub and Iur interfaces, as well as user data from the packet-switched services on the Iu interface, is carried by AAL5.

The ATM layer brings together functions connected with handling ATM cells, i.e. adding the header, multiplexing cells on the physical layer, and routing of cells (the ATM layer is found at the level of the ATM network switches).

Figure 6.10 shows a summary of the different operations carried out by the different layers of the ATM.

The AAL/CS layer adds a header to the user data, and possibly (depending on the type of AAL) an end of message marker including an error correcting code. The end of the frame may contain padding bytes in order to ensure that the frame can be split into a whole number of ATM cells. The AAL/SAR layer cuts the frame into blocks of 48 bytes. The cell header is then added by the ATM layer.

6.5.2 Transport Protocols

As a general rule, signalling exchanges on the control plane use ATM/AAL5 on all the UTRAN interfaces (Iu, Iub and Iur). User plane data itself is carried by

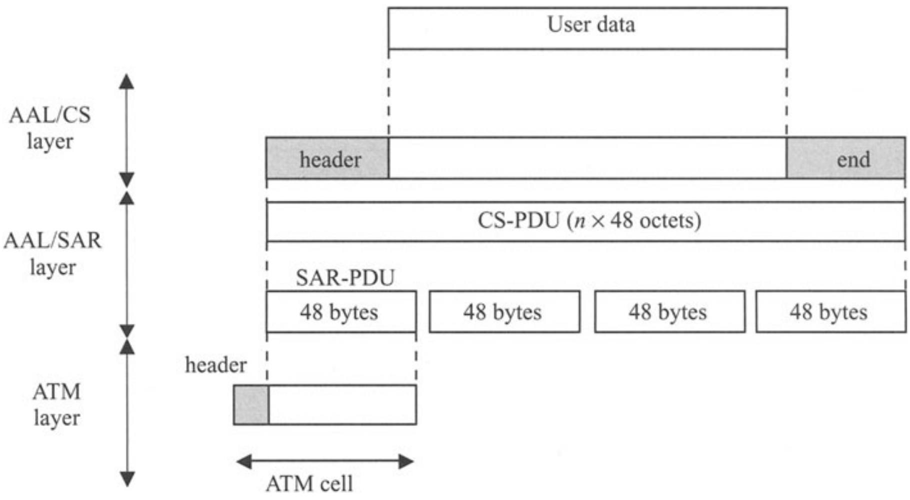


Figure 6.10 Functions of the ATM layers.

ATM/AAL5 if it belongs to the packet-switched domain and by ATM/AAL2 if it belongs to the circuit-switched domain (Figure 6.11).

6.5.3 Transport Layer Signalling (ALCAP)

The access link control application part (ALCAP) is used by the transport layer to set up transmission paths in the user plane (Figure 6.12).

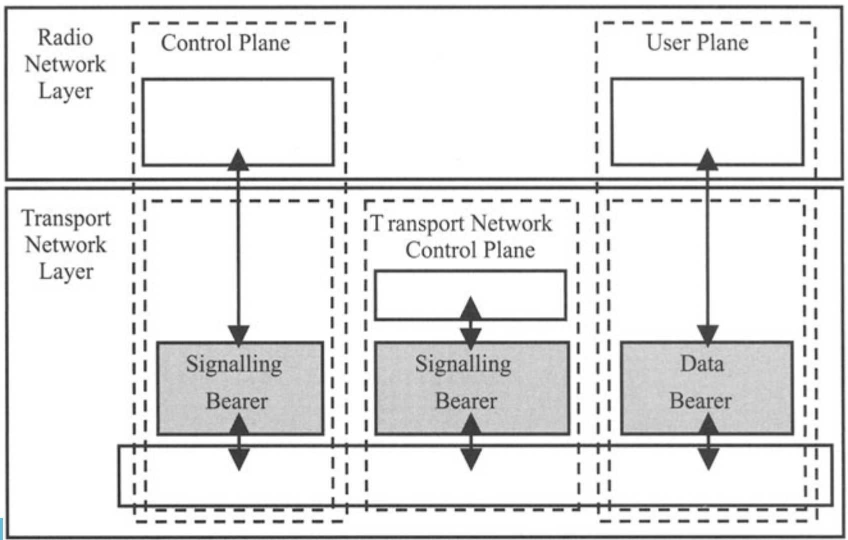


Figure 6.11 The transport layer.

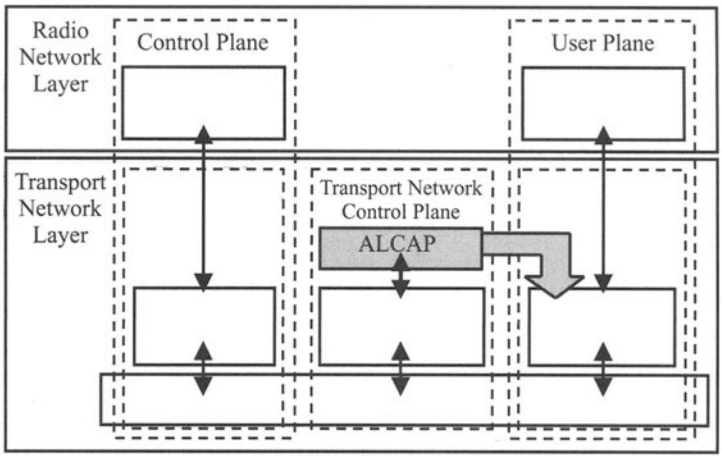


Figure 6.12 Transport layer signalling.

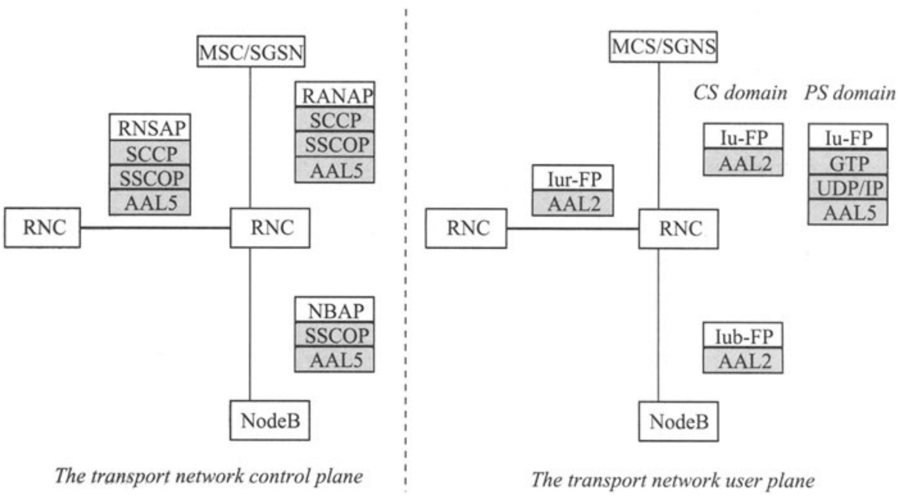


Figure 6.13 The UTRAN network interface layers.

When the user plane uses ATM/AAL5 transport, the virtual circuits are pre-established. In this case neither the ALCAP nor the transport network control plane exist. When the user plane employs ATM/AAL2 transport, the virtual circuits are set up through the ALCAP, which, in the case of the AAL2 signalling protocol, is defined by the Q.2630.1 standard.

6.6 Summary

Figure 6.13 provides a summary view of the layers used on the UTRAN network interfaces. The grey blocks correspond to the layers belonging to the transport networks.

The service-specific connection-oriented protocol, used above the AAL5 transport layer, allows the transport of signalling information to be made reliable, thanks to error correction and detection functions and to integrity control of the information transported.

7 Call Handling

This chapter describes the functions used by the mobile and the network during the setting-up of a call. In order to present a more global picture of the mechanisms in play, it also addresses the operations carried out when the terminal is switched on or off and when it registers with the network.

7.1 Switching On

As soon as it is switched on, the mobile carries out a certain number of operations aimed at selecting a network and a cell able to offer a service to the user, i.e. capable of sending a call to the network or receiving one from it. These operations are grouped into three processes within the mobile:

- selection of a public land mobile network (PLMN);
- selection of a cell;
- registering with the network.

7.1.1 Concept of PLMN

The PLMN is defined in the standard as a telecommunication network made up of a core network and an access network, installed and managed by an operator.

In general, each country has several GSM PLMNs (in the UK, there are four: Cellnet, Vodafone, T-Mobile and Orange). When UMTS is deployed, it is very likely that new operators (and therefore new UMTS PLMNs) will see the light of day.

Each PLMN has an identity made up of two fields (Figure 7.1):

- the *mobile country code* (MCC) identifies the country in which the PLMN is located. All the PLMNs in the same country have the same MCC;

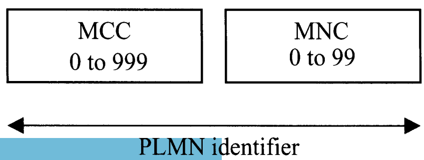


Figure 7.1 PLMN identifier fields.

Table 7.1 Examples of PLMN codes.

Country	MCC	MNC	Operator
United Kingdom	234	10	Cellnet
		15	Vodafone
		30	T-Mobile
		33	Orange
Germany	262	01	T Mobil
		02	D2
		03	Eplus
France	208	01	Itinéris
		10	SFR
		20	Bouygues Télécom
Italy	222	10	Omnitel
		01	TIM

- the *mobile network code* (MNC) identifies the different PLMNs within the same country.

These two fields form an integral part of the *international mobile subscriber identity*, which identifies each subscriber who has subscribed to a PLMN operator.

The MCC is allocated by an international body, the International Telecommunications Union (ITU), in order to avoid the same code being used by two different countries. In contrast, the allocation of the MNC is left to the regulatory bodies of each country. By way of illustration, Table 7.1 contains examples of the PLMN codes for a number of European operators.

The PLMN is not restricted to subscribers who have an account with the operator of the network; it is also available to users whose accounts are with other operators, provided that an agreement has been reached between the operator of the network visited and the operator of the network with which the user holds an account. In this case, the user is said to be *roaming*.

7.1.2 Concept of Equivalent PLMNs

Development of the Operator's Role

In GSM networks and, more generally, in most second generation networks, the service provider also owns and operates the network infrastructure. The operators of networks identified by different PLMN codes therefore use separate infrastructures (Figure 7.2).

This one-to-one correspondence between service provider and infrastructure operator was defined when work on the GSM standard was first started. It imposes a particular way of working on the mobile and on the fixed network that makes any other commercial model difficult to implement. However, the use of auctions for allocating third generation network licences, and the extremely high prices reached in bidding for them, have helped to bring the operators together, because they see co-operation as a way of maximising the chances of getting a licence and of

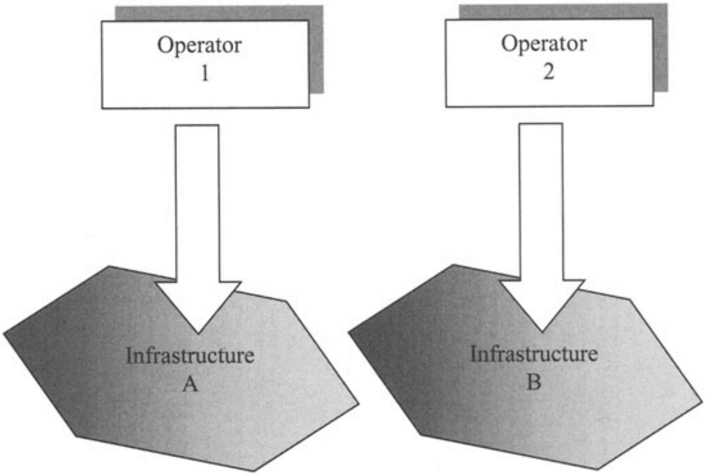


Figure 7.2 Second generation networks.

reducing the cost of deploying and managing the network infrastructure. Further, in some countries, changes in the regulatory regime are allowing a new type of operator to appear, the *mobile virtual network operator* (MVNO).

MVNOs have the special feature that they offer a service in association with a network operator who owns all or part of the infrastructure as well as the licence to use part of the radio spectrum. There thus exist different scenarios, in which an MVNO may own a smaller or larger part of the infrastructure, for example, the core network, part of the access network, or only the home location register. As a minimum, an MVNO must hold the details of the SIM/USIM cards of its subscribers, since these are necessary for identifying the user, making the exchanges secure, and granting access to the services offered.

The emergence of the third generation has therefore raised important questions about the roles of network operators and service providers as they are defined in the GSM standard.

The following paragraphs describe some examples of relations between infrastructure operators and service providers that might possibly come into existence when third generation networks are operational.

One Infrastructure Shared by Several Service Providers

There exist at least three scenarios in which service providers, possibly competing, might be led to share the use of a single network infrastructure (Figure 7.3).

- Two UMTS operators each having a UMTS licence decide to share the UMTS access network infrastructure in less densely populated (and hence less profitable) areas but to develop separate infrastructures in the more densely populated areas. This type of strategy allows subscribers to both networks to be provided with national coverage, while reducing the costs of creating and maintaining the infrastructure.

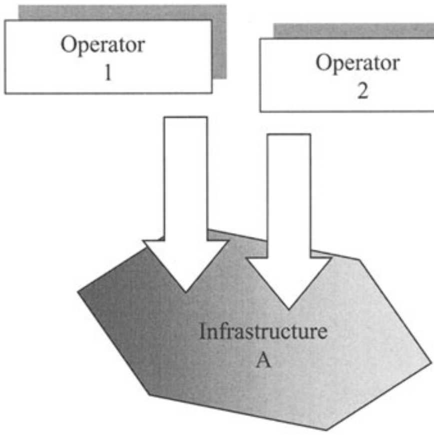


Figure 7.3 Two service providers sharing a network infrastructure.

- Two GSM operators decide to exploit a UMTS licence jointly. This scenario could occur, for example, when two GSM operators decide to share the expenses of acquiring a licence, or following failure in the bidding process.
- If we postulate a complete separation between the activities of managing a network and providing a service, it could happen that two MVNOs might offer competing services based on the same network infrastructure, run by a third party.

Two Different Networks Run by a Single Operator

It may happen, because of the constraints of national regulation, that a GSM operator who has acquired a UMTS licence is forced to use different PLMN identities. In this case, mobiles in idle mode will consider the two networks to be different (Figure 7.4).

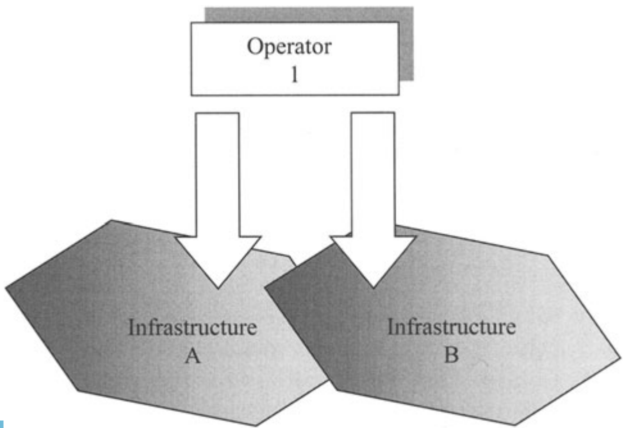


Figure 7.4 Two sets of infrastructure used by a single operator.

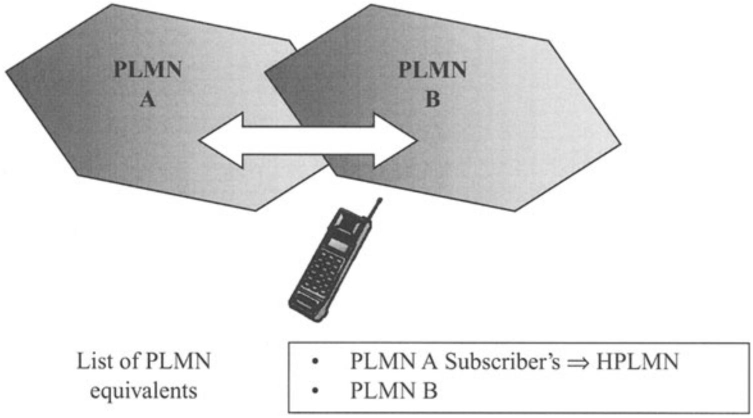


Figure 7.5 An example of PLMN equivalence.

Conclusion

The concept of the PLMN, common to the GSM and UMTS standards, will have to be developed so that it can offer more flexibility in network deployment.

Equivalent PLMNs

In order to respond to third generation network needs as far as the independence of the network operator and the service provider are concerned, the concept of 'equivalent PLMNs' was introduced into the GSM and UMTS standards at the beginning of 2000 (Figure 7.5).

The principle is as follows. The USIM (or SIM) of the user is not changed and remains attached to the operator with whom the user has an account. On the other hand, the network can indicate to the mobile, either when it is registering with the network or when it changes localisation zone, the PLMNs that it must consider as equivalent when reselecting a cell in idle mode. The mobile can thus pass from one PLMN to another as if they formed a single network.

The list of equivalent PLMNs is held in the memory of the mobile and is erased when the mobile is switched on, which gives the network operator the opportunity of changing the list, as a result of new or modified agreements with other operators, without having to update all its clients' USIM cards.

7.1.3 Selecting the PLMN

When it is switched on, the mobile searches for accessible PLMNs by reading the beacon channels of the surrounding cells. The mobile carries out this search by looking at all the possible frequencies for the access technologies supported. (The process used by the mobile to search the UTRAN FDD cells was described in Section 5.3.1.)

When a GSM or UTRAN/FDD beacon channel is identified, the mobile reads the system information broadcast on this channel in order to obtain the identifier of the PLMN to which the cell is attached.

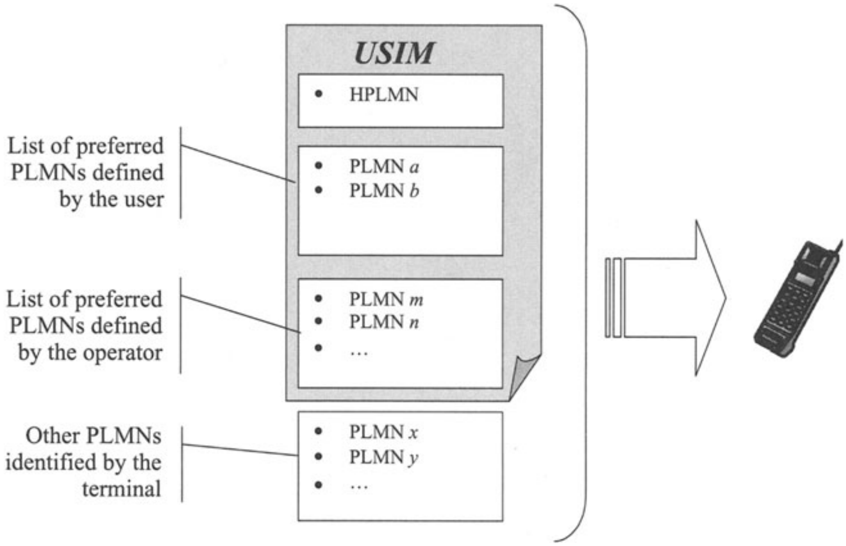


Figure 7.6 Priority classification of the PLMNs accessible to the terminal.

Figure 7.6 describes the way in which the terminal classifies the accessible PLMNs.

In the first place, if the user's home PLMN (i.e. the PLMN of the operator with whom the user has an account) is accessible, the mobile will try to register with it.

If the mobile is not in an area covered by its home PLMN, the mobile searches for the accessible PLMNs in a preferential list defined by the user and stored on the USIM card. If this fails, or if the list has been left empty by the user, the mobile carries out a similar search using the preferential list defined by the operator. If the result is still negative, the mobile will select the PLMN that corresponds to the most favourable cell from the radio transmission point of view.

Of course, as is the case with GSM, the user can always select a PLMN manually from among those detected by the mobile, in order to try registering with it. If this is successful, the user can then initiate or receive calls from another user or from the fixed RTCP network in the PLMN visited.

Searching for a PLMN of Higher Priority

Subsequently, while in the idle state (i.e. during the periods in which no service is active), the mobile will try periodically to find a PLMN of higher priority than the one chosen when it was switched on. The frequency with which this operation is carried out is fixed by the operator and stored in the USIM card. It is not modifiable by the user. Its value lies between 6 minutes and 8 hours, with a default value of 60 minutes.

This mechanism allows operators to be sure that, when their users are roaming nationally, they will be registered again with their home PLMN when this becomes accessible. It also allows users who are roaming and have defined a list of preferred

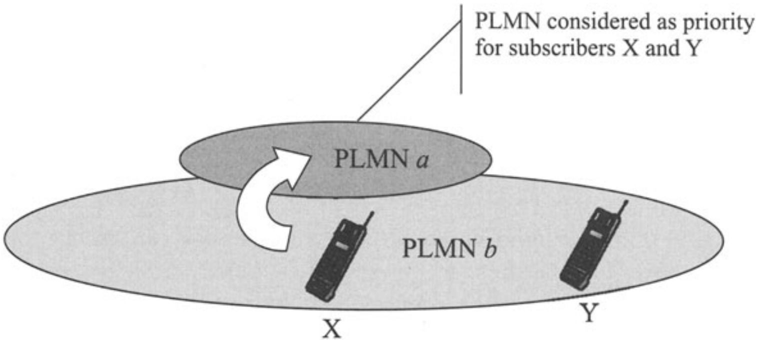


Figure 7.7 Selection of a higher priority PLMN.

PLMNs (for example, for tariff reasons or the availability of services) to use their choice of networks when this is available.

This mechanism is illustrated in Figure 7.7. Subscribers X and Y are roaming in PLMN *b*. In the geographical area covered by both PLMN *a* and PLMN *b*, user X will select PLMN *a*, at the end of a search period.

The mechanism for searching for a PLMN of higher priority is used in all UMTS mobiles and in recent GSM ones (i.e. ones that are compatible with version 8 of the GSM specifications, published in 1999). In preceding versions of the GSM standard, the mechanism was limited to searching for the home PLMN when roaming nationally (when the MCC field of the PLMN visited has the same value as that of the home PLMN). The extension introduced thanks to UMTS means that the behaviour of mobiles in idle mode is better adapted to the different grouping patterns of the operators or to the sharing of infrastructure referred to earlier.

7.1.4 Searching for Candidate Cells

Before selecting the cell to register with, the mobile must first create a list of suitable candidate cells. A cell is considered suitable if it satisfies certain conditions:

- It must not be barred. This information is broadcast on the cell's beacon channel.
- It must satisfy the S radio criterion.

The S Criterion

The S criterion is defined by the following expression:

$$(S_{\mathrm{qual}} > 0) \textbf{ and } (S_{\mathrm{rxlev}} > 0)$$

S_{qual} is a measure of the quality of the cell, defined by

$$S_{\mathrm{qual}} = Q_{\mathrm{qualmeas}} - Q_{\mathrm{qualmin}}$$

where Q_{qualmeas} is the ratio E_b/N_o for the common pilot channel of the cell (which allows the quality of the cell to be estimated) and Q_{qualmin} is the minimum level of quality required.

S_{rxlev} is a measure of the signal level received from the cell. It is defined by:

$$S_{rxlev} = Q_{rxlevmeas} - Q_{rxlevmin} - P_{compensation}$$

where $Q_{rxlevmeas}$ is the signal level received by the mobile on the common pilot channel, $Q_{rxlevmin}$ is the minimum acceptable level of the signal, and

$$P_{compensation} = \max\{P_{max\_RACH} - P_{max}, 0\},$$

P_{max_RACH} being the maximum power level that the mobile can use on the RACH and P_{max} the maximum transmission power of the mobile. The P term measures the power reserve of the mobile on the uplink. When the maximum transmission power of the mobile is below the level permitted in the cell, the received level of the signal, S_{rxlev}, is reduced.

This means that when the signal level received by the mobile is weak (i.e. ($Q_{rxlevmeas} - Q_{rxlevmin}$) is only slightly larger than zero), the cell can be considered acceptable ($S_{rxlev} > 0$) if the transmission level is compensated by sufficient transmission power on the uplink (i.e. $P_{compensation} = 0$).

The S criterion thus allows the mobile to set up a list of cells that satisfy the requirements of quality and minimum power at the same time.

Choosing the Initial Cell

Having set up a list of candidate cells, the mobile must then select one cell from the list to try to register with. To do this, it classifies the cells by some radio criterion (either the ratio E_b/N_o for the common pilot channel or the level of signal received on that channel) and then chooses the best cell according to this classification.

7.1.5 Registering with the Network

Once the cell and the PLMN have been selected, the mobile tries to register with the chosen PLMN for each of the service domains offered by the core network, i.e. the CS and PS domains. In the UMTS specifications, this procedure carries the name:

- *IMSI attach* for registering with the CS domain;
- *UMTS GPRS attach* for registering with the PS domain.

If there is no roaming agreement between the home PLMN operator and the operators of the PLMNs that the mobile can see, the registration process fails. In this case the mobile is allowed to initiate emergency calls only.

Figures 7.8 and 7.9 show the signalling exchanges between the network nodes concerned with the registration procedure. The principal protocols employed in the procedure are the following:

- GPRS mobility management (GMM), between the mobile and the SGSN, for registering with the PS domain;
- the MM (mobility management) protocol, between the mobile and the MSC/VLR, for registering with the CS domain;

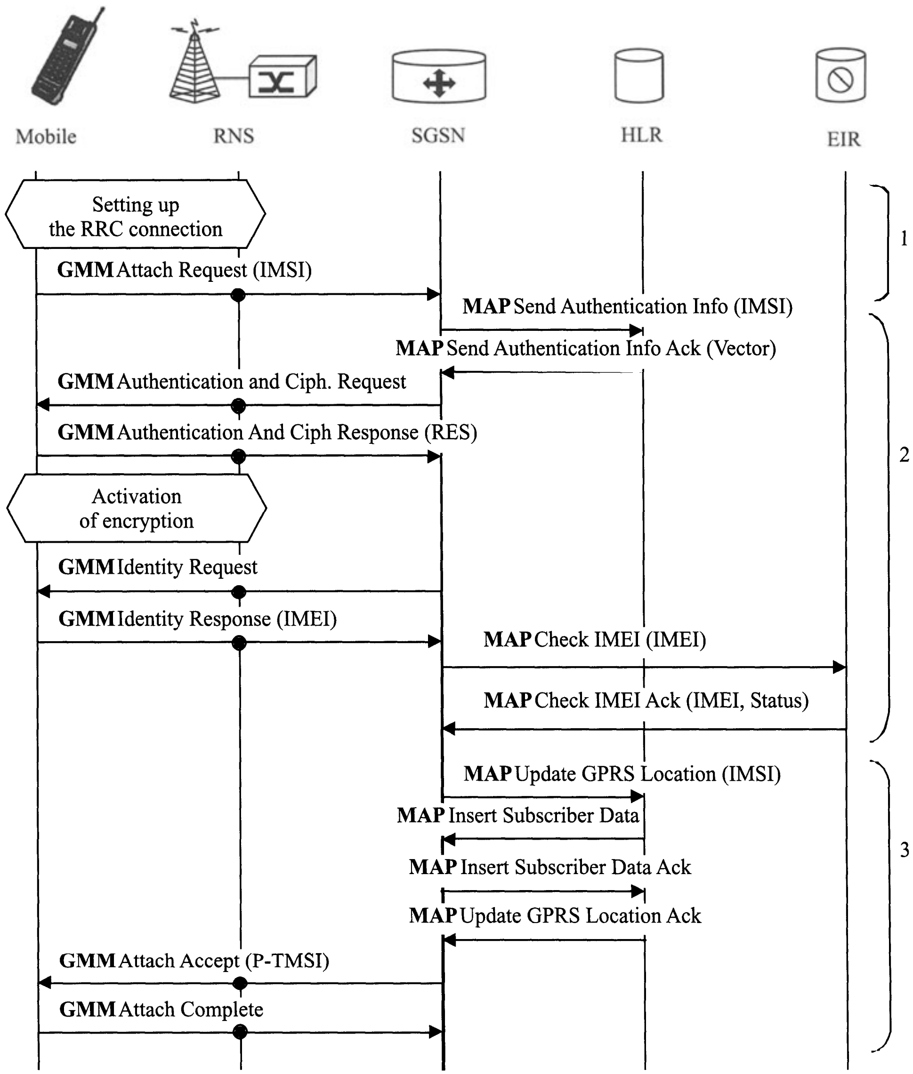


Figure 7.8 Registering with the PS domain.

- the MAP (mobile application part) between the different nodes of the core network.

Registering with the PS Domain

The process of registration is made up of three main phases:

1. Once the RRC connection is set up between the mobile and the RNC, the registration request is sent to the SGSN by the mobile via the RNS.

2. Before it can register the mobile, the SGSN must carry out certain checks regarding the validity of the user's identifier (and, more precisely, of the IMSI stored on the SIM card) and the identity of the mobile.

The user's identity is checked through an authentication procedure. The data that allows this authentication to be carried out (the authentication vector), is first requested from the HLR by the SGSN.

Checking the terminal identity is an optional procedure. At the request of the SGSN, the terminal provides its identity (IMEI: International Mobile Equipment Identity). The EIR, interrogated by the SGSN, responds with a message saying whether or not the terminal is on the black list.

During this phase, encryption over the UTRAN interfaces is switched on, in order to protect later signalling exchanges between the mobile and the network, in particular the allocation of a temporary identity. The authentication procedure as well as other security-linked functions are described in more detail in the section on call set-up.

3. Once the identities have been checked, the SGSN can proceed to register the terminal with the network. The SGSN informs the HLR that the mobile is registered in its database. In response, the HLR transmits to the SGSN the parameters of the user's account. This information will be used later by the SGSN when the user wants to initiate a data transfer to or from the network.

The last operation carried out is the allocation of a temporary identity (P-TMSI: Packet Temporary Mobile Station Identifier). It is this identity that will be used in later exchanges between the mobile and the network.

The process of registering with the CS service domain is almost identical to that for the PS domain, even to the names of the messages. On completion of the process, an identity specific to the CS domain is allocated to the module: a temporary mobile station identity (TMSI).

Combined CS/PS Registration

In order to reduce the number of signalling exchanges between the mobile and the network, the standard offers the option of registering with CS and PS services simultaneously, provided the network supports the optional Gs interface between the SGSN and the MSC/VLR. The initial phase of the combined procedure is identical to that for PS. Once the mobile is registered in the SGSN database, the SGSN sends a request to the MSC/VLR for the mobile to be registered. The MSC/VLR informs the SGSN of the success of the CS registration by the Location Update Message, which contains the temporary CS identity of the user (TMSI). At the end of the procedure, the mobile is thus allocated the two temporary identities TMSI and P-TMSI, one for each domain (Figure 7.9).

Registration Failures

It may happen that the mobile's request to register is refused by the network. Figure 7.10 show the different failure cases that can occur.

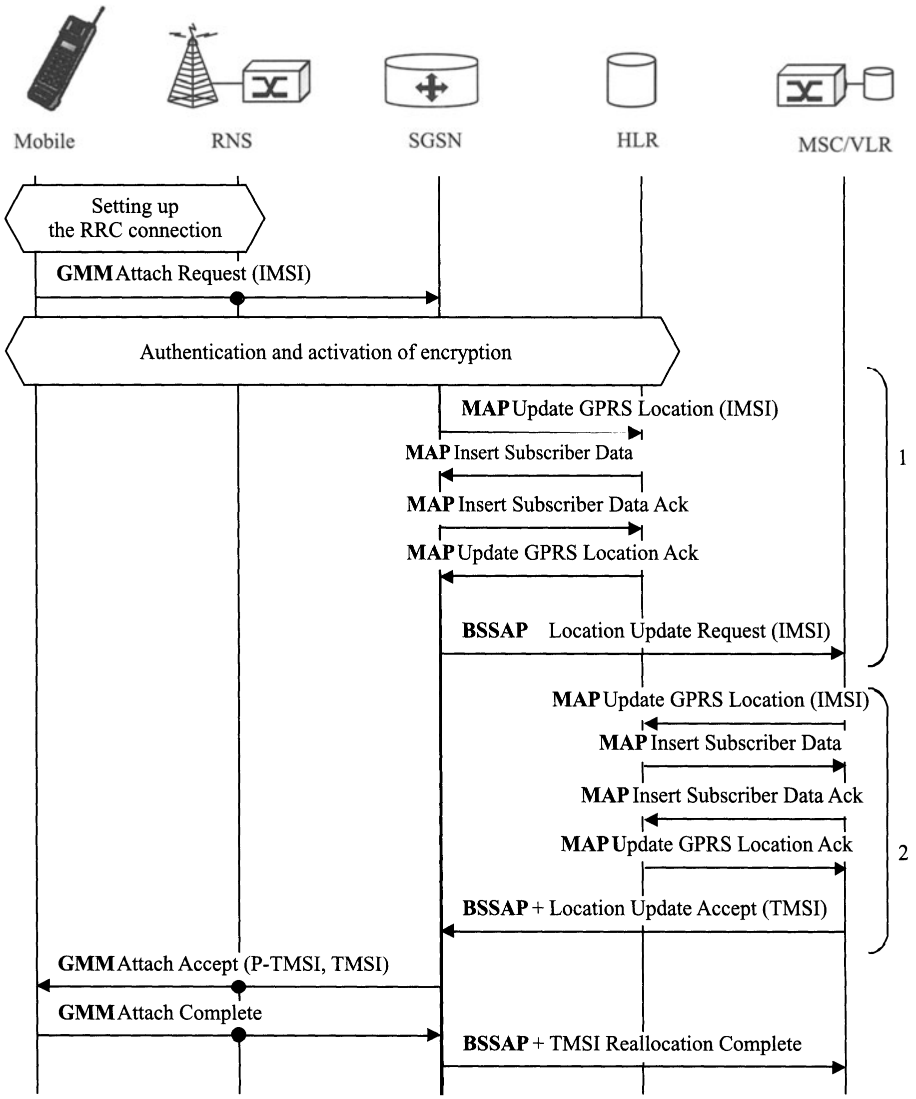


Figure 7.9 Combined CS/PS registration.

If the registration request comes from a stolen terminal or SIM card, whose theft has been noted by the operator, the request will be refused either by the HLR (in the case of an unknown or invalid IMSI) or by the EIR (if the IMEI has been put on the black list). The USIM card is then declared invalid by the mobile, which will refuse to try to register again until it is replaced.

There are more common situations that can lead to a registration request being rejected, for example, there may be no roaming agreement between the subscriber's home operator and the operator of the network being visited, or the subscriber may not be authorised to change network. In these cases, the identity of the forbidden

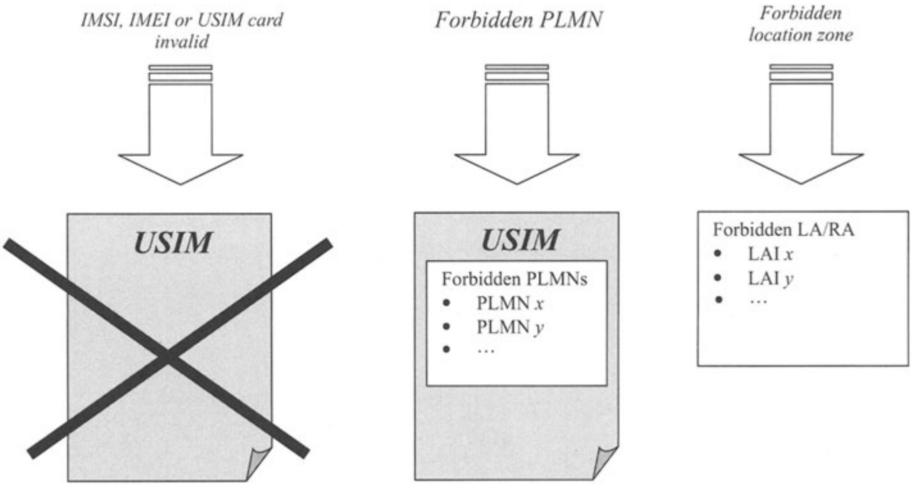


Figure 7.10 Network registration failure.

PLMN is stored in the USIM card, thus allowing the information to be retained when the mobile is switched off.

The operator also has the option of locally restricting access to the network by rejecting registration requests in certain zones. This practice, for example, allows part of the network resources to be kept for certain categories of user. In this case, the identity of the forbidden zone is stored in the mobile. The mobile can try to register again when a candidate cell belonging to another zone is detected.

7.1.6 Terminal Display

The information displayed on the terminal in idle mode is of particular importance to the operators. It is the means by which users (their own users as well as roaming users) are informed about service availability. For users, it is the most direct way of knowing what network they are on and, in particular, whether or not they are within the area covered by their home PLMN. Depending on this, the services available and the charging rates can be very different.

The procedure employed in the GSM networks to display the name of the network operator is extremely simple. Each terminal holds in ROM a table of correspondences between PLMN codes and the name of the network or operator. This is the name that is displayed on completion of the registration procedure.

This technique has several weaknesses. In particular, when a change of trading name occurs, the name displayed by the mobile cannot be modified. In the scenarios described above, with operator mergers or the sharing of infrastructure, this lack of flexibility could carry quite a severe penalty.

In the UMTS standard, this mechanism has been retained but improved. UMTS mobiles still have a fixed table of correspondences. The operator, however, has the means, as shown in Figure 7.11, to erase this information if it becomes out of date.

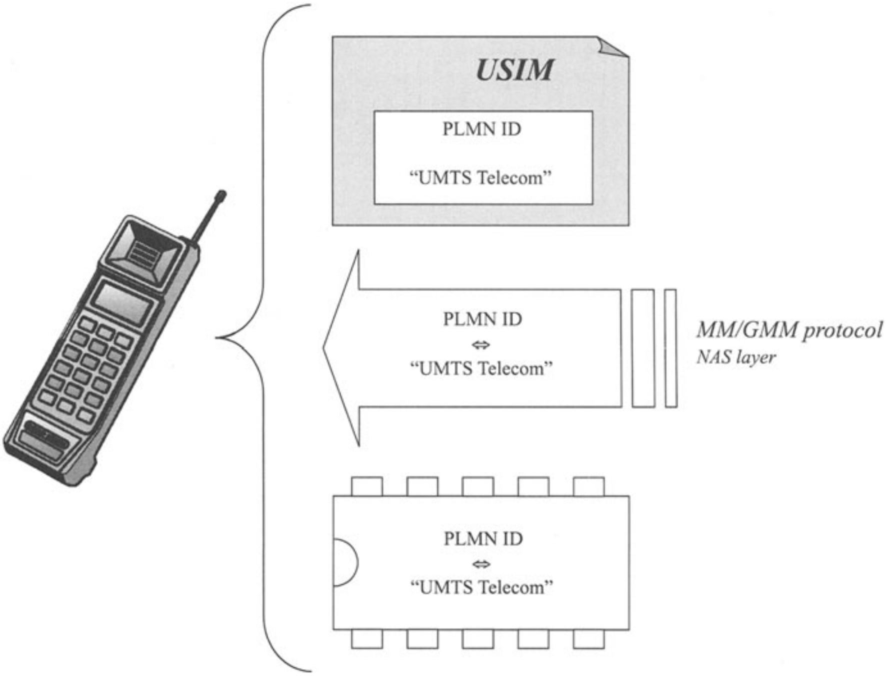


Figure 7.11 Displaying the network name.

First of all, operators can store their own table of PLMN/operator name correspondences on the USIM card. Even if it remains difficult to change the content of the USIM cards already in service, at least operators can make sure that those of new subscribers are up to date.

Secondly, at the end of the registration procedure, the operator can provide the mobile with the name of the network with which it is registered. This information is sent to the mobile by means of the INFORMATION message, which forms part of the MM and GMM protocol layers.

The UMTS standard defines an order of priority among these three sources of information:

- If the USIM card contains a name for the current PLMN, this is the name that must be displayed.
- Next in order of priority comes the name provided during the MM/GMM signalling exchanges.
- Finally, if no other information about the current PLMN is available, the mobile displays the name stored in its ROM.

7.1.7 General View

Figure 7.12 brings together the different processes that are triggered when the mobile is powered up.

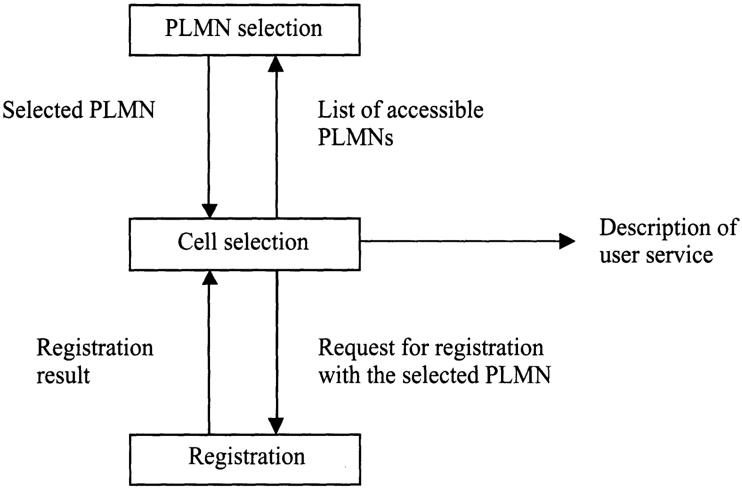


Figure 7.12 Switching on the terminal.

The process of cell selection looks for the PLMNs that are accessible on the frequency bands that correspond to the access technologies supported by the mobile, i.e. for a mobile supporting both GSM and FDD technologies, the 900 MHz and 1800 MHz GSM bands and the UTRAN/FDD band

The selection process then goes on to choose which PLMN to try to register with. The PLMN chosen is:

- the home PLMN when this is available;
- another PLMN, chosen according to the priority rules, possibly specified by the USIM card.

The process then tries to register with the best of the candidate cells belonging to the chosen PLMN and satisfying criterion S.

Depending on the result of the registration procedure, the user is sent an indication of the service available, either:

- the name of the PLMN with which the mobile is registered is displayed on the screen; or
- if no PLMN has accepted the registration request, the mobile indicates to the user that only emergency calls can be made.

7.2 Call Set-up

This section describes the various phases and procedures used during the setting up of PS or CS calls. In order to understand the call diagrams, it is useful to be familiar with the following ideas:

- the PDP context;

- the various states of the mobile;
- security procedures.

7.2.1 PDP Context

The packet data protocol (PDP) context is peculiar to the PS domain. It collects together all the information that allows user data to be transmitted between the mobile, the UMTS network and the external packet-switched network of the Internet.

The PDP context (Figure 7.13) contains the following main data items:

- the quality of service associated with the call, which is in fact represented by the attributes of the radio access bearer allocated by the UTRAN;
- the access point name, which is the identifier of the external PDP network that the mobile wants to access;
- the PDP address of the mobile. In the case of an external Internet network, this will be an IPv4 or IPv6 address.

The attributes of the PDP context are the values belonging to each user, depending on the nature of their account. One might thus imagine that access to certain PDP networks providing particular services would be limited to a subset of the users of the UMTS network.

Before initiating any data transfer in PS mode, the mobile must ask the network to activate a PDP context; the network will then have to check that the attributes of the context asked for are compatible with the user's account details. The call set-up diagrams shown in this section show the different phases in the activation of a PDP context.

The user can activate several sessions in parallel, for example to collect mail simultaneously from two electronic mailboxes maintained by two different service providers. In such a case, the user must activate as many PDP contexts as sessions.

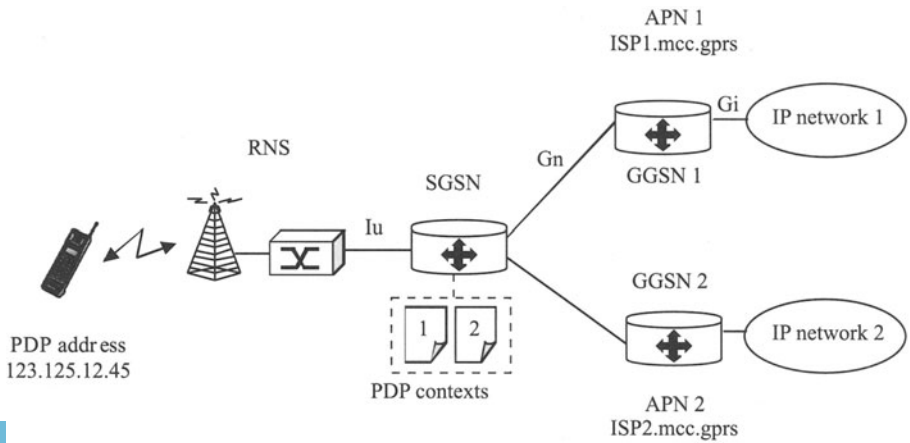


Figure 7.13 The PDP context.

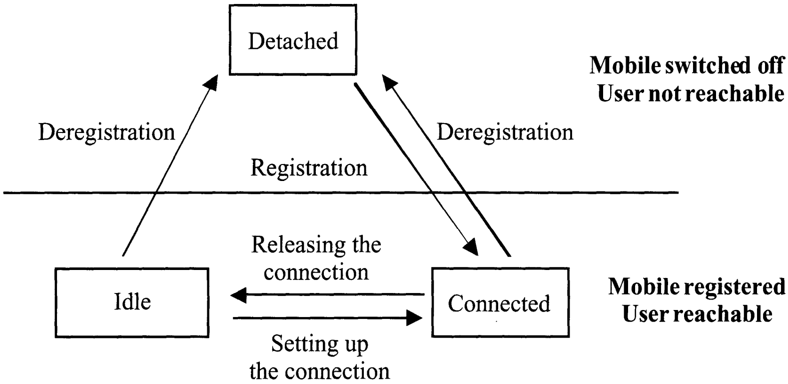


Figure 7.14 The MM states of a mobile in circuit-switched mode.

7.2.2 States of the Mobile

In CS mode, the mobile can be in any one of three states: when the mobile is switched off, it is said to be in the *detached* state; once the mobile is switched on and registered with a PLMN, as described in the previous section, it is said to be in the *idle* state if no service is active; finally, when communication is in progress, the mobile is said to be in the *connected* state (Figure 7.14).

When the mobile registers with the CS domain, user mobility is handled differently depending on the state of the mobile. In idle mode, mobility is under the control of the mobile through the cell re-selection procedure. In connected mode, mobility is under the control of the network, through the handover procedure. Mobility functions are addressed in more detail in Chapter 8.

In PS mode, the GMM states of the mobile are identical, which represents a change from the position for the GPRS service on GSM networks (Figure 7.15).

In GPRS the connected state does not exist. Once the mobile registers in GPRS mode, upward or downward transfers of packets are carried out through sessions

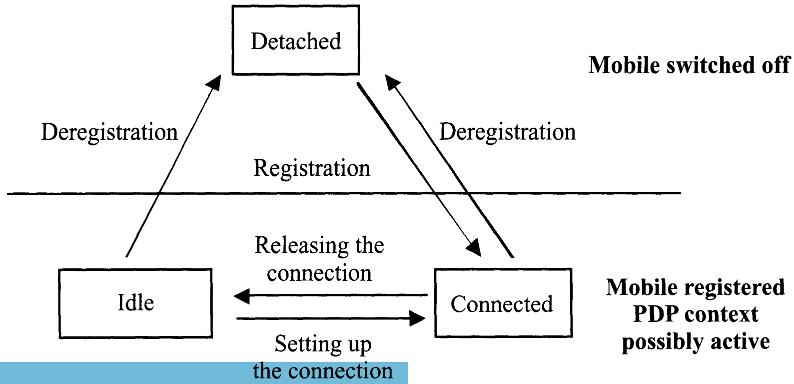


Figure 7.15 The PMM states of a mobile in packet-switched mode.

called *temporary block flows*, which are set up and then closed down when the transfer is terminated. For example, in the case of a Web navigation application, the transfer block flow is active only for the time it takes to download an HTML page.

UMTS networks work differently. We have seen that the RRC layer of the UTRAN was introduced in order to maintain a connection between the mobile and the network even when no data was being transferred, while not using up any radio interface resource. While in this state, user mobility is managed by the serving RNC.

Despite everything, it is possible in UMTS to free the connection when passing into the idle state. The mobile remains reachable from the network provided it has a PDP address, whether permanent (in which case it forms part of the subscriber's account details held by the HLR) or provided by the network during the activation of the PDP context. In the idle state, the mobility of the user is handled at the SGSN level by the routing area update procedure, described in more detail in Chapter 8.

7.2.3 Security Procedures

The three security procedures proposed in the UMTS standard and described in this section are authentication, encryption and integrity checking.

Authentication

Authentication plays a fundamental role in the security of UMTS networks. The procedure has three objectives. First of all it allows the network to check the identity of the subscriber, the purpose being to prevent mobiles with a forged USIM from accessing the network (or, at the very least, to make it very difficult). Conversely, the procedure also allows the mobile to authenticate the network, by checking the validity of the information it transmits.

Finally, the authentication procedure also allows the mobile to generate the IK and CK encryption and integrity checking keys, using a mechanism described in the next section.

Authentication rests on certain items that must be kept secret:

- the key K, which is specific to the user. It is known only by the USIM card and the Authentication Centre (AuC), a database situated in the core network;
- the F1, F2 and F5 algorithms, whose use is described later.

Figure 7.16 shows the stages in the authentication procedure carried out by the MSC/VLR for the CS domain. In the case of the PS domain, the procedure is carried out by the SGSN but follows an identical path, almost to the names of the messages.

The authentication procedure is carried out by the network each time the mobile accesses it, be it for the initial registration with the PLMN, for a response to a paging message, or for a call initiated by the user of the mobile.

When a request for service is received from a user, the MSC/VLR asks the HLR/AuC for the information needed to authenticate the user, identified by his IMSI.

Starting from the key, K, associated with the IMSI, a random parameter, RAND, and the F2 algorithm, the HLR/AuC produces the expected response code (XRES)

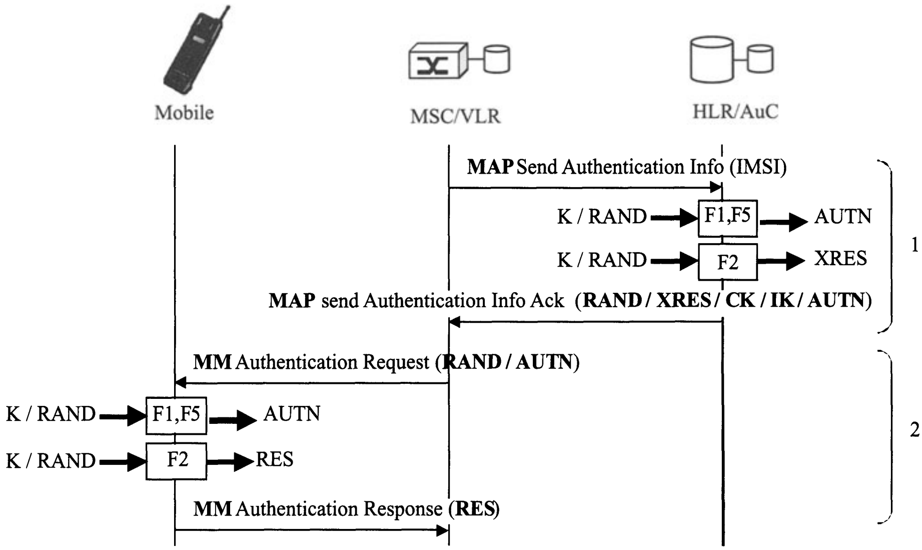


Figure 7.16 The authentication procedure.

that will allow the user's USIM card to be authenticated. The response of the HLR/AuC consists of the following five items, known as the authentication vector:

- the random parameter RAND;
- the code XRES;
- the key IK;
- the key CK (the way in which the IK and CK keys are generated and used will be described in the next section);
- the authentication token, generated by the F1 and F5 algorithms, which will then be used by the mobile to authenticate the network.

The SGSN then sends the mobile a request for authentication, containing the RAND parameter and the authentication token. The mobile verifies the latter using the F1 and F5 algorithms. If the verification fails, the request for authentication is rejected by the mobile, which returns to the idle state. The mobile must then consider the current cell blocked, i.e. it will not try to access the network again using this cell.

If the network is correctly authenticated by the mobile, it will send the network a response code, RES, calculated from the key, K, contained in the USIM card, and the parameter, RAND, using the F2 algorithm. The RES and XRES codes are then compared with each other by the SGSN. In case of failure, the mobile is informed that the authentication has been rejected. The mobile then considers the USIM card contained in it to be invalid until the next time the mobile is switched on or a new card is inserted.

Integrity Checking and Encryption

Encryption allows the confidentiality of user data exchanged between the mobile and the network, in either direction, to be guaranteed. Encryption is applied independently to the PS and CS domains of the core network and concerns all the signalling and user data associated with each domain.

Integrity checking, a new mechanism not in GSM, is applied to the exchange of signalling information between the mobile and the network. It allows the receiving entity to authenticate the sender and to assure itself that the message received has not been forged or altered in the course of transmission. Integrity checking allows all the RRC signalling messages between the mobile and the SRNC to be protected, as well as the exchanges between the mobile and the core network, corresponding to the MM and CC layers (mobility management and call control for the CS domain) and the GMM and SM layers (for the PS domain). Only a few RRC messages, such as the messages used for setting up a connection or the messages for broadcasting system information, are not protected by the integrity checking mechanism.

Like encryption, integrity checking is applied independently to the PS and CS domains of the core network and concerns all the signalling associated with each domain.

Unlike the encryption mechanism offered by the GSM standard, the UTRAN encryption, like the integrity checking, is implemented by the layers of the UTRAN radio protocols situated in the RNC in charge of the RRC connection with the mobile. Thus encryption and integrity checking protect not only transmission on the radio interface but also the Iub network interface, and possibly the Iur when a drift RNC is in use.

In order to improve the robustness of these mechanisms, we will see that a minimum of configuration information must travel between the mobile and the network. Other items necessary for encryption and integrity checking derive from algorithms kept secret and known only to the mobile and certain nodes of the network.

The encryption and integrity checking rest on the use of a specific key for each mechanism (Figure 7.17):

- IK (integrity checking key);
- CK (ciphering key).

These keys are produced in the mobile and in the HLR/AuC, starting from the following items:

- a random parameter, RAND, given to the mobile during the authentication phase;
- a secret key, K, known only to the core network and the USIM card;
- the F3 and F4 algorithms, which are also kept confidential.

Data encryption is carried out by the RLC layer when the RLC protocol is used in acknowledged or unacknowledged mode, and by the MAC layer when the RLC is used in transparent mode (Figure 7.18).

Each MAC or RLC block is simply encrypted by adding bit by bit (with no carry, i.e. an exclusive OR operation) the bits of the data block and the bits of the

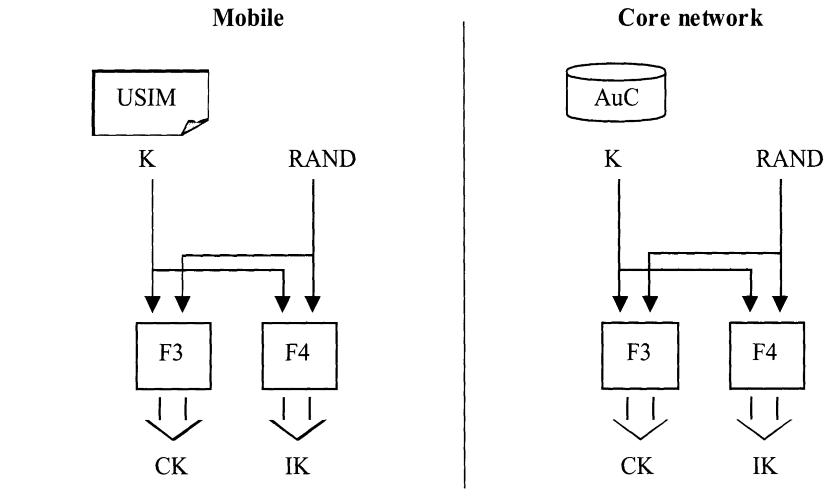


Figure 7.17 Generating keys for encryption and integrity checking.

keystream block (KSB). The KSB itself is generated by the F8 algorithm, which uses the following parameters:

- the encryption key, CK;
- a counter, COUNT-C, which is a sequence number that varies over the course of time;
- a length parameter, LENGTH, which allows the F8 algorithm to generate KSB blocks of the same length as the data blocks to be encrypted;

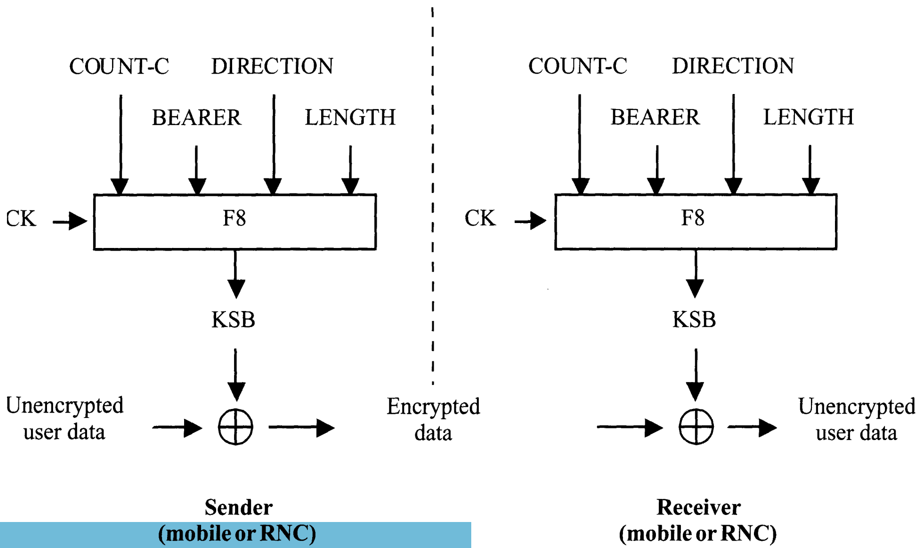


Figure 7.18 Encryption.

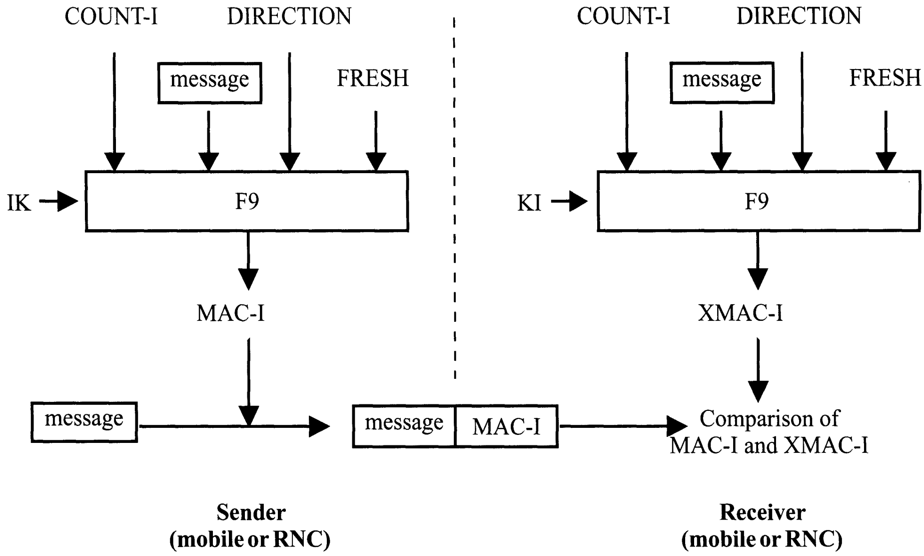


Figure 7.19 Integrity checking.

- the parameters, DIRECTION and BEARER, used to avoid the possibility of different transport channels having the same KSB, which would weaken the protection provided.

The receiver carries out exactly the same operations to decrypt the blocks received. An identical KSB is generated, then added bit by bit to the bits of the encrypted block to recover the user data.

Integrity checking is carried out in a very similar manner (Figure 7.19).

The F9 algorithm generates a message authentication code (MAC) starting from the message to be sent, the integrity checking key, IK, and parameters COUNT-I and DIRECTION. The parameter FRESH is used to strengthen the protection provided. FRESH is a random value exchanged between the network and the mobile when the RRC connection is set up; it guarantees to the network that the mobile is not reusing old message authentication codes.

The MAC is then concatenated with the message before transmission. On receiving the message, the receiver generates a code XMAC (expected MAC) using the same method as the sender and then compares it with the MAC received. If this check fails, the message is considered invalid and is destroyed by the receiver.

Activation of Encryption and Integrity Checking

When activating encryption and integrity checking, it is important to follow the correct sequence of operations, otherwise there is a risk, for example, that one of the two entities receives an encrypted (and therefore unintelligible) message when it is expecting to receive a message in clear (i.e. not encrypted).

Activation of encryption and integrity checking always takes place at the request of the core network; more precisely, the requests are made by the MSC for circuit-switched services and by the SGSN for packet-switched services, through the security mode command message (Figure 7.20). This request is then relayed to the mobile using the RRC message of the same name, this message being sent in RLC acknowledged mode.

Activation of integrity checking is relatively simple. The RNC applies integrity checking to messages sent by, and received from, the mobile as soon as it receives confirmation that the security mode command message has been correctly received by the mobile.

In the case of encryption, activation is not immediate. The network specifies a starting time to the mobile, after which data must be transmitted in encrypted mode. This starting time is expressed in the form of a level 2 RLC or MAC frame number, according as to whether the channel is in RLC transparent mode or in acknowledged or unacknowledged mode. (We saw in the chapter on the UTRAN radio interface that encryption is carried out at the level of the MAC layer for a channel in transparent mode and at the RLC level for a channel in acknowledged or unacknowledged mode.) When the frame number is reached, both the mobile and the network activate encryption mode for sending and receiving. Figure 7.20 gives a general picture of the activation of encryption and integrity checking for a single communication.

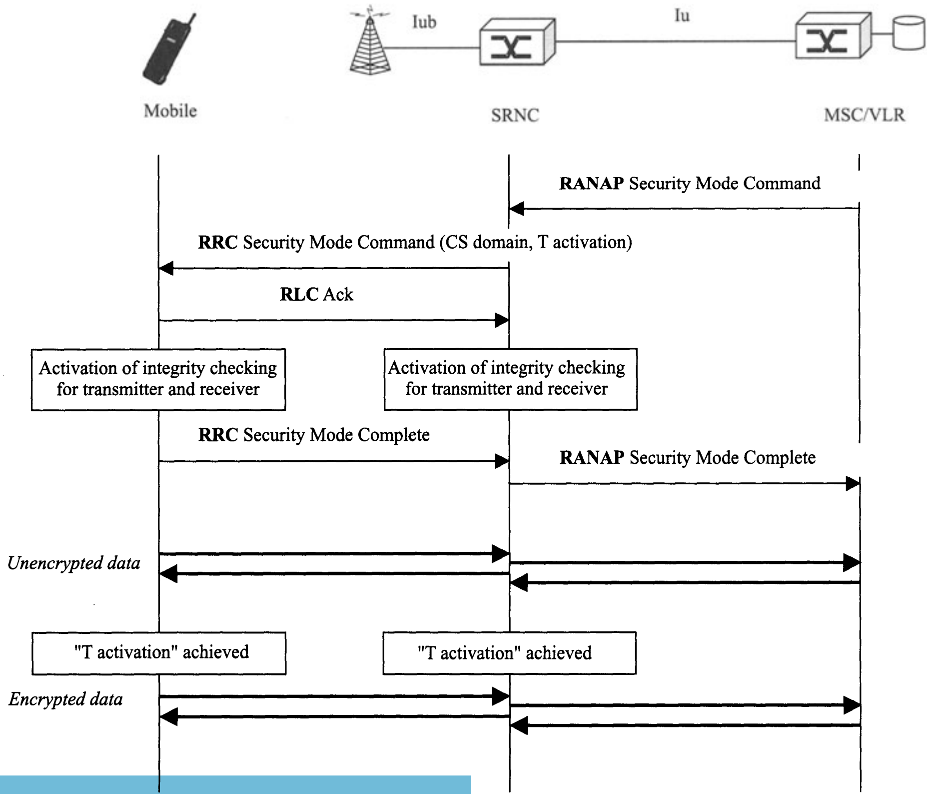


Figure 7.20 The transition to encrypted mode.

7.2.4 Network Access

In the chapter on the physical layer, we saw the structure of the physical random access channel (PRACH) and how it is used in accessing the network. In this section we describe the mechanisms available in the UMTS that the operator may use to regulate and control access to the network (for example, in case of traffic congestion).

Access Classes

The idea of access class was introduced into the standard to regulate or limit access to the network in case of exceptional circumstances. The idea had been introduced in GSM and was taken up in an identical manner in the UMTS specifications.

Network subscribers are all divided randomly into ten access classes (numbered 0 to 9), described as ordinary or public. In addition, certain subscribers may also belong to one or more special classes, depending on the nature of their activity. The standard defines five special classes (numbered 11 to 15), as follows:

- class 11, reserved for the network operator;
- class 12: security services (police, surveillance, etc.);
- class 13: public services (electricity distribution, etc.);
- class 14: reserved for the emergency services;
- class 15: reserved for the staff of the network operator (allocated, for example, to on-site maintenance technicians).

Access class 10 corresponds to emergency calls, i.e. calls to special numbers, 999 in the UK, 911 in North America, and 112 in continental Europe.

The network transmits continuously, on the beacon channel BCCH, the access classes authorised to access the channel. Thus, when a mobile wants to access the network (whether at the request of the user or on receipt of a paging message), it will compare the access class or classes allocated to the user and stored in the USIM card, with the classes authorised by the network.

Operators thus have a very simple means of limiting new calls to their networks. By successively forbidding public access classes (each of which represents 10% of the population of subscribers), they can go as far as completely forbidding access to their networks. This type of access control can be very relevant in the case of traffic congestion or in the case of natural disaster, in order to reserve communication resources for emergency calls and the emergency services.

Service Classes

As we saw in Chapter 5, the PRACH is a contention channel, just as in GSM. The number of collisions increases with the number of users wanting to access the network, which increases the number of access failures, in a way that is independent of the service required by the user, be it an emergency call or a simple text message.

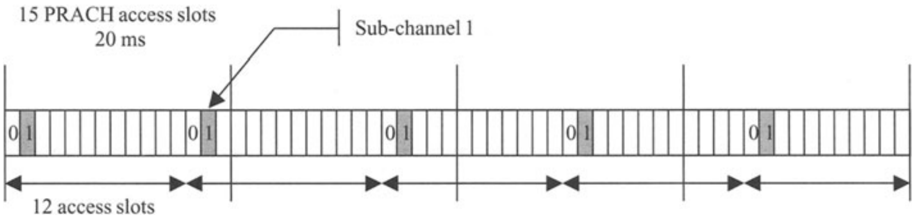


Figure 7.21 Dividing the PRACH into 12 sub-channels.

To alleviate this difficulty, the UMTS standard defines a resource sharing mechanism for the PRACH, based on access service classes (ASC). The access capability of the PRACH can be presented as a set of 16 signatures usable on a set of 12 sub-channels. As shown in Figure 7.21, a sub-channel of the PRACH is a subset of PRACH access slots. The distance between two consecutive access slots of the same sub-channel is 12 access slots. The operator can thus partition the RACH capacity by allocating to each partition (or service class) a subset of the signatures or sub-channels of the PRACH.

The relation between access classes and service classes is then broadcast over the beacon channel and read by the mobiles situated in the cell. Each mobile in the cell thus knows the class of PRACH service that it can use, depending on the access class of the user as stored in the USIM.

Figure 7.22 shows an example of the partitioning of the access capacity of the PRACH. ASC 0 is accessible to all users of the network, because it is associated with access classes 0 to 9. ASC 1 is reserved for the special access classes, 11 to 15. ASC 2 is superimposed on the previous two. It represents the access capacity of the PRACH for emergency calls.

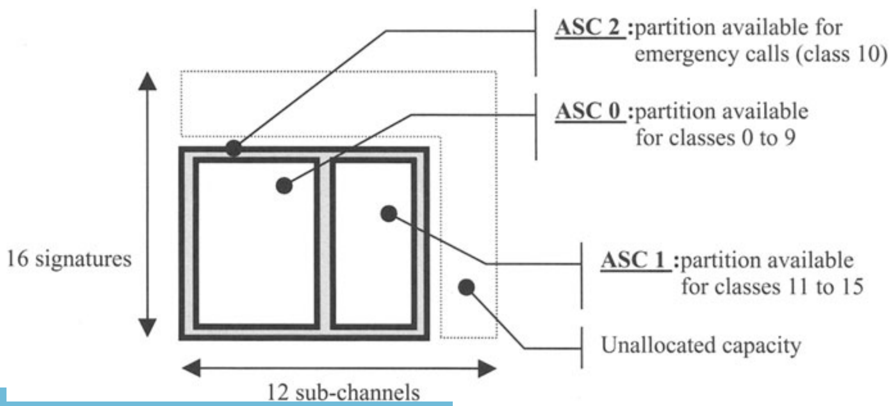


Figure 7.22 An example of the partitioning of the PRACH.

7.2.5 Setting Up a PS Call

A PS call corresponds to the activation of a PDP context by the mobile, whether at the request of the mobile or of the network. The example in Figure 7.23 shows the setting up of a PDP context at the request of the mobile.

1 The first operation carried out is the establishment of an RRC connection between the mobile and the network. Once this is done, the mobile transmits the initial message, *activate PDP context*, to the network, along with the domain for which it is intended (the PS domain in our example).
   Before transmitting this message to the core network, the SRNC sets up the SCCP connection with the SGSN.
2 The network then carries out various operations intended to authenticate the user (the GMM *Authentication and Ciphering Request*) and make the transmission secure (*Security Mode Command*).
3 The PDP context request can finally be handled by the network. If the parameters of the context required by the mobile are compatible with the subscriber's account details, the SGSN will order the setting up of all the core network and UTRAN resources needed for the call.

At the level of the core network, the tunnel between the SGSN and the GGSN is set up by means of the procedure *Create PDP Context*.

At the level of the UTRAN, the allocation of resources is ordered by the message *RAB Assignment Request*. The SRNC must then:

- configure the Node B according to the radio resources chosen (the NBAP procedure of *Radio Link Setup*);
- tell the mobile the parameters of the radio resources allocated (the RRC procedure of *Radio Bearer Setup*);
- set up the AAL2 virtual circuit or circuits on the Iub interface and the GTP tunnel on the Iu interface, corresponding to the radio resources allocated (AAL2 signalling messages *Establish Request* and *Establish Confirm*).

When the allocation phase is completed, the SGSN indicates to the mobile that the PDP context has been created. It is then possible to transfer user data.

7.2.6 Deactivation of the PDP Context

Both the user and the network can deactivate the PDP context or contexts that have been allocated. In the case of deactivation requested by the mobile, the mobile sends its network request (the *Deactivate PDP Context Request*) to the SGSN, which will delete the contexts allocated and send the deactivation request to the GGSN.

7.2.7 Setting Up a CS Call

Outgoing Calls

Figure 7.24 shows an example of setting up an outgoing call, i.e. a call initiated by the mobile.

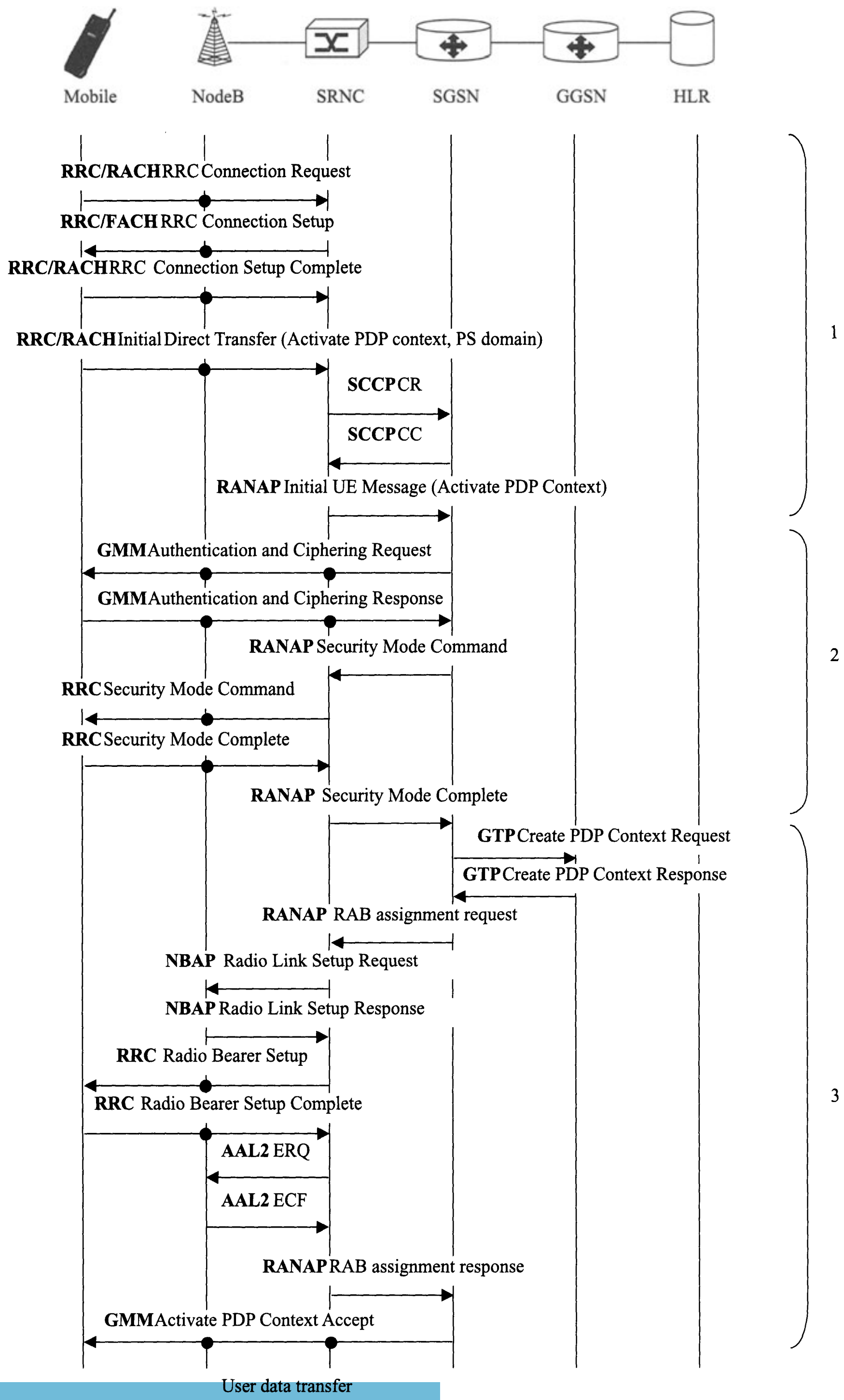


Figure 7.23 Setting up a PS call.

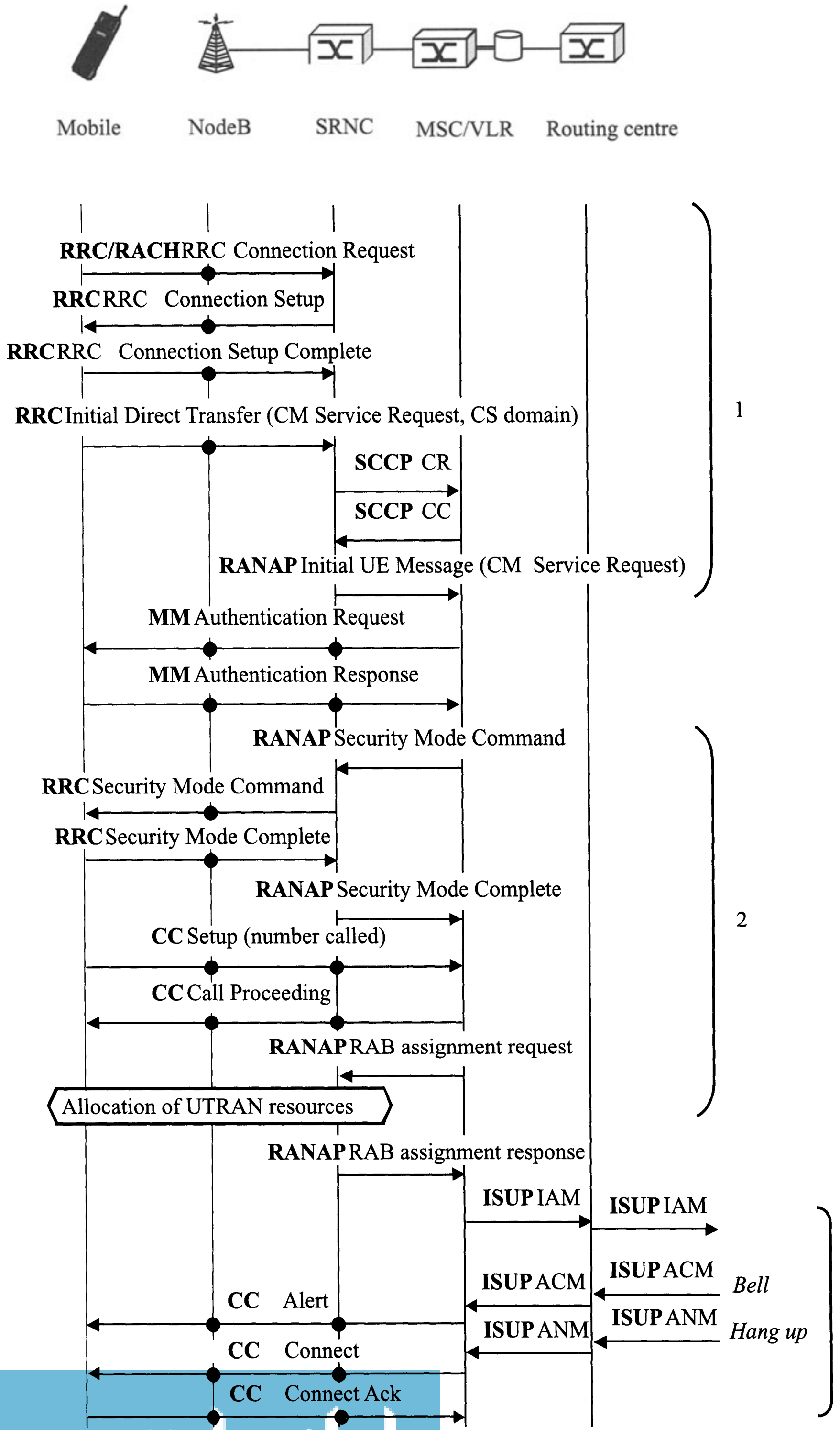


Figure 7.24 Setting up an outgoing call.

Setting up the RRC connection is similar to the PS case; the only difference lies in the level of the initial message transmitted to the MSC/VLR.

1 Before accepting the call, the MSC/VLR must carry out the authentication and possibly shift to encrypted mode.
2 Completion of the encryption procedure indicates to the mobile that the request for service has been accepted by the network. The mobile then sends the *setup* message containing, amongst other things, the number to be called. The *call proceeding* message allows the network to signal to the mobile that it has all the necessary items available to go ahead and handle the call.

   The network will then allocate the necessary resources in the UTRAN, through the RAB assignment procedure.
3 Once the resources are allocated, the call will be routed to its destination via the routing centre, situated outside the UMTS network, by means of the *Initial Address Message* (IAM) of the ISUP layer, containing the number called.

   The *Address Complete Message* indicates that the called number has been alerted. This information is relayed to the caller by the MSC/VLR.

   The *Answer Message* indicates that the number called has responded and the call is established.

Incoming Calls

An incoming call is a call to a mobile user, initiated by another mobile subscriber or by a fixed subscriber. The example in Figure 7.25 shows the different phases in setting up an incoming call when the caller is a fixed subscriber.

1 When the caller from the fixed network has finished dialling the number of the mobile subscriber being called, the call is routed to one of the MSCs of the mobile subscriber's PLMN. This MSC will play the role of the gateway MSC (GMSC).

   The GMSC has no information about the position of the mobile subscriber in the network. It will therefore interrogate the HLR, which in turn interrogates the VLR in which the subscriber is located, to obtain a mobile station routing number (MSRN), a number that will be used in routing the call from the GMSC to the destination MSC.

   In our example, the subscriber's mobile is switched on, it has carried out the registration procedure and is in a location area of its home PLMN. The call will therefore be routed by the GMSC to the MSC/VLR covering the current location area of the mobile.

   On receiving the IAM, the MSC/VLR will ask for a paging message to be broadcast to the current location area of the subscriber. Broadcasting of this message into the different cells of the location zone is done by the RNC or RNCs in charge of the zone.

   For reasons of security, the MSC/VLR uses the temporary identity of the mobile (the TMSI) when paging.
2 The rest of the progress of the call is identical to that for an incoming call. The RRC connection is set up, the initial message is routed to the MSC/VLR (in this

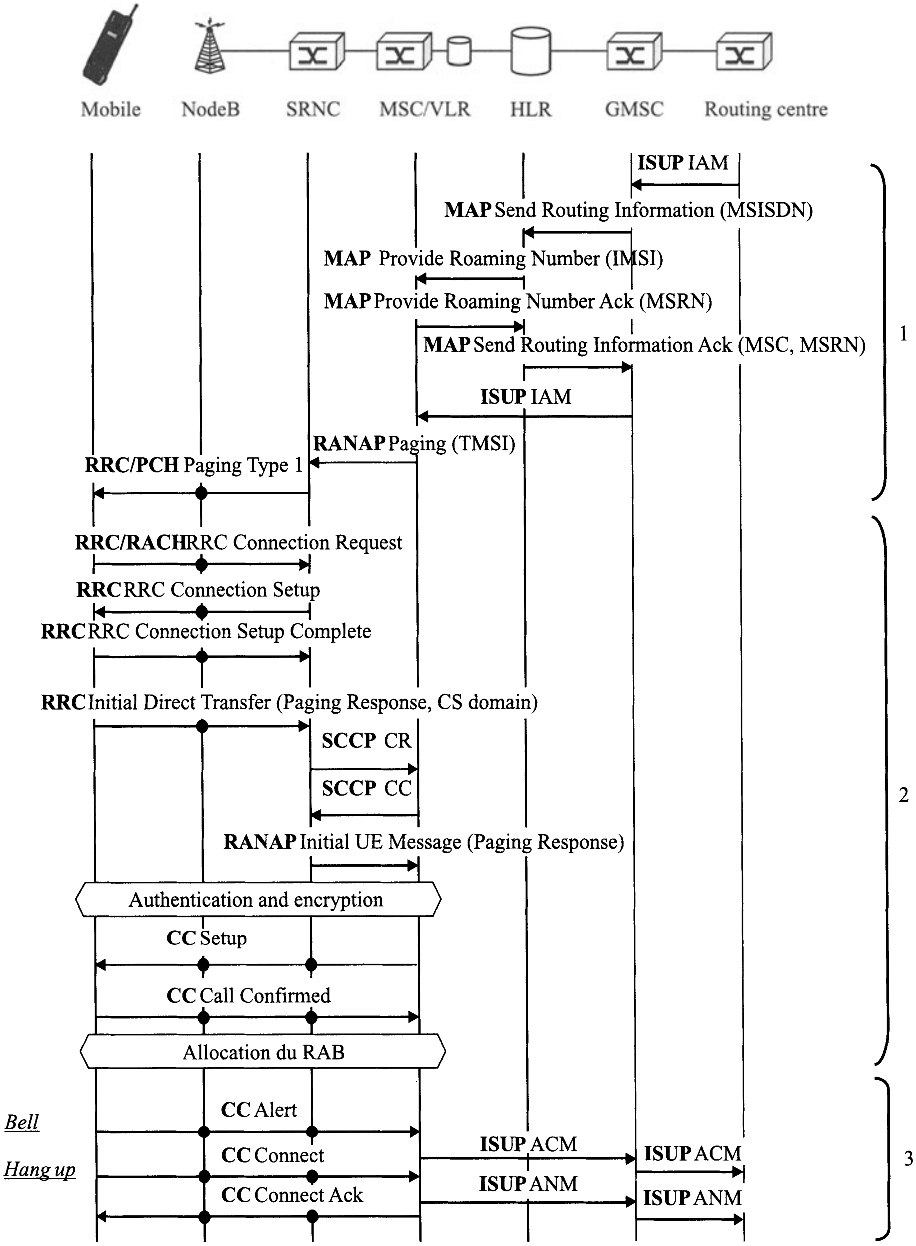


Figure 7.25 Setting up an incoming call.

case the message is *paging response*), the mobile is authenticated and the necessary resources are allocated.

3 When the mobile's bell rings, the *alert* message is sent to the MSC and relayed to the caller on the fixed network by the *Address Complete* message.

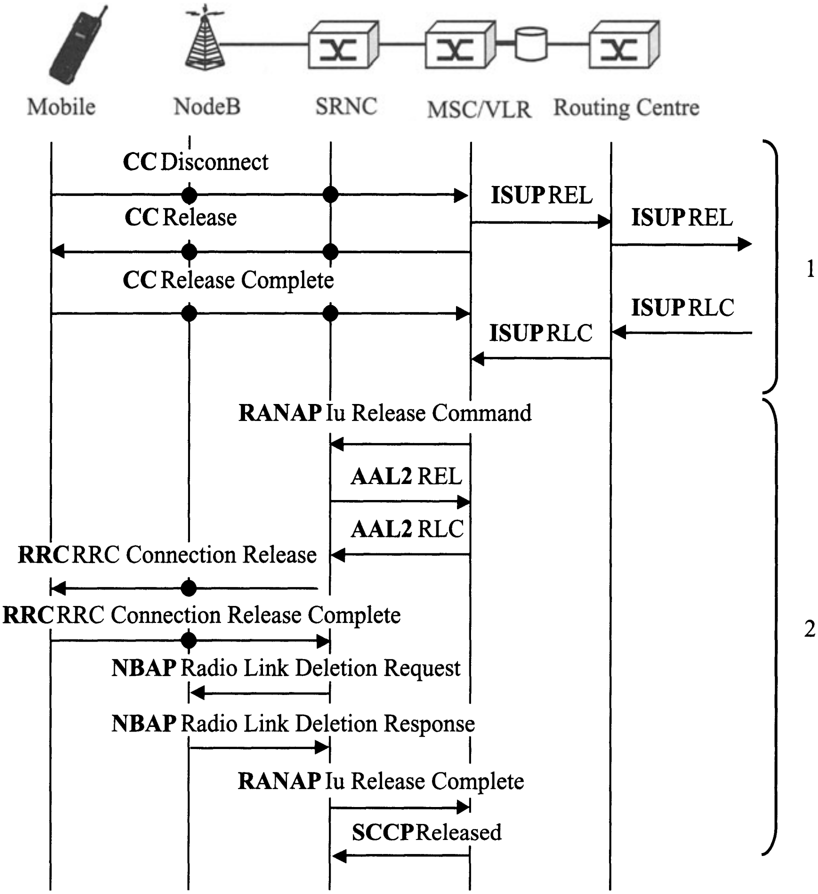


Figure 7.26 Call termination initiated by the called party.

7.2.8 Terminating a CS Call

Figure 7.26 shows the end of a CS call initiated by the mobile.

1 When the user hangs up, the message *disconnect* is sent to the MSC/VLR. The end of call indication is transmitted to the fixed network by the ISUP message *release* (REL).

2 Once the call has been terminated, it remains necessary to free the resources used in the UTRAN, i.e. the AAL2 circuit (freed by the ALCAP AAL2 exchange *release/release complete*), the RRC connection, the radio link or links, and the SCCP connection between the SRNC and the MSC/VLR.

7.3 Switching Off

When the mobile is switched off, the *IMSI detach* procedure (or *GPRS detach* in the case of the PS domain) may be used by the mobile to delete its registration with the network. This procedure is not obligatory; its use is controlled by the network.

The purpose of the *IMSI detach* procedure is to delete the location data linked to the subscriber in the VLR and to inform the HLR of the users deregistration.

The network operator may wish not to allow the use of the *IMSI detach* procedure if he prefers not to support the signalling load it generates. Since each VLR has a database cleaning procedure, the mobile will in any case be deregistered from the network after a certain time, which is of the order of the length of time it takes to update the location area. This point is also addressed in Chapter 8.

7.4 Service Billing

Billing for services is part of a much larger set of mechanisms for managing network events. These mechanisms include, in particular:

- billing for services activated by users of the network, whether they have an account directly with the operator or are working through the intermediary of a service provider or are roaming;
- statistics about the operation of the network (number of calls per hour, mean call length, etc.). Detailed analysis of the call profiles of network users enables the operator to size the network better or to anticipate peak loadings;
- management of call records, used, for example, when a user challenges the size of his bill.

The list of possible operations authorised by the standard is very large indeed. The purpose of this section is simply to describe event management applied to service billing, an operation important for the user and critical for the network operator since it is directly linked to his turnover.

7.4.1 Network Architecture

Service billing is a function carried out by network equipment quite separate from that which we have addressed so far. Figure 7.27 shows the architecture of a UMTS network, completed by:

- the billing centre (BC). The BC receives call detail records from the various nodes of the network, which will allow the service used to be billed;
- the charging gateway function (CGF), which is a gateway between the SGSN/GGSN and the billing centre. The CGF, which belongs to the core network PS domain, is used as an intermediate storage point for call detail records, several records for the same user possibly being issued by different nodes of the network.

This architecture is completely independent of the access network and is therefore just as applicable to a UMTS network as to a GSM network. Note also that the billing centre is linked, directly or indirectly, to two types of network node: the MSC and SGSN on the one hand, and the GMSC and the GGSN on the other.

The MSC and the SGSN, as network nodes directly connected to the access network, are responsible for collecting information relating to the access network. These reports always contain the identity of the subscriber (IMSI) and possibly the

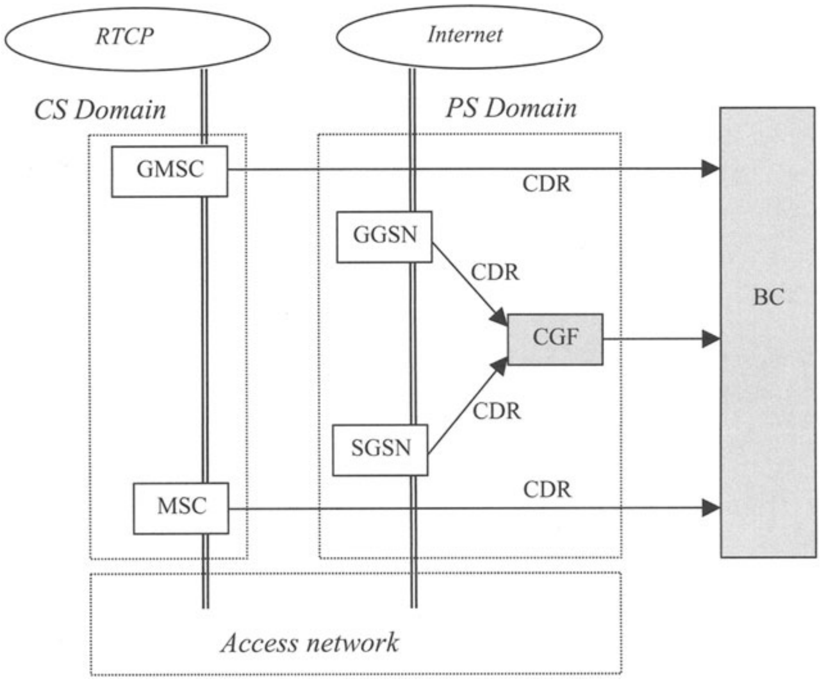


Figure 7.27 Network architecture and billing.

identity of the mobile (IMEI). In the case of the MSC, they can also contain the effective data rate used on the radio interface or the AMR mode used for a telephony application.

The GMSC and the GGSN, as gateways to external networks (the Internet or the PSTN), collect information concerning the second segment of the call. There is some duplication between the reports sent by the MSC and the GMSC, for example, the call duration or the number of data packets sent. This duplication in fact allows the information directly linked to billing to be checked for consistency and to be safeguarded against loss.

In principle, the standard fixes no rules about how calls are billed. In practice, a certain number of rules seem to be commonly accepted and followed by most operators.

For example, for a circuit-switched call, normal usage is to charge the user according to the length of time that the resource is used. For a packet-switched call, in contrast, the charge will generally be dependent on the number of bytes transmitted, since the access network can free the resources allocated to the user during periods of inactivity (this is the nominal function of GPRS networks). In the case of UMTS, freeing of resources means that the network has to carry out supplementary operations, such as handling the CELL_FACH state.

It is also accepted that a caller trying to reach a mobile subscriber must be charged a sum that is independent of the actual location of the mobile. In particular, if the mobile subscriber is outside his home PLMN, the extra to be paid on the call (arising from the routing of the call between the two networks) should be charged to the mobile subscriber.

On the other hand, although the standard allows the possibility, no operator directly charges for the registration process (*IMSI attach*) nor for updating the location area (*Routing Area Update* and *Location Area Update*). These procedures nonetheless generate signalling traffic in the network that is by no means negligible and are activated in very different ways, depending on individual subscribers' habits.

7.4.2 Some Examples

This section describes two examples illustrating the billing mechanisms:

- a packet-switched call from a mobile to an information server;
- an incoming, circuit-switched call (e.g. a telephone call), with the mobile subscriber located on another PLMN.

Accessing an Internet Server

Figure 7.28 illustrated the progress of this type of call.

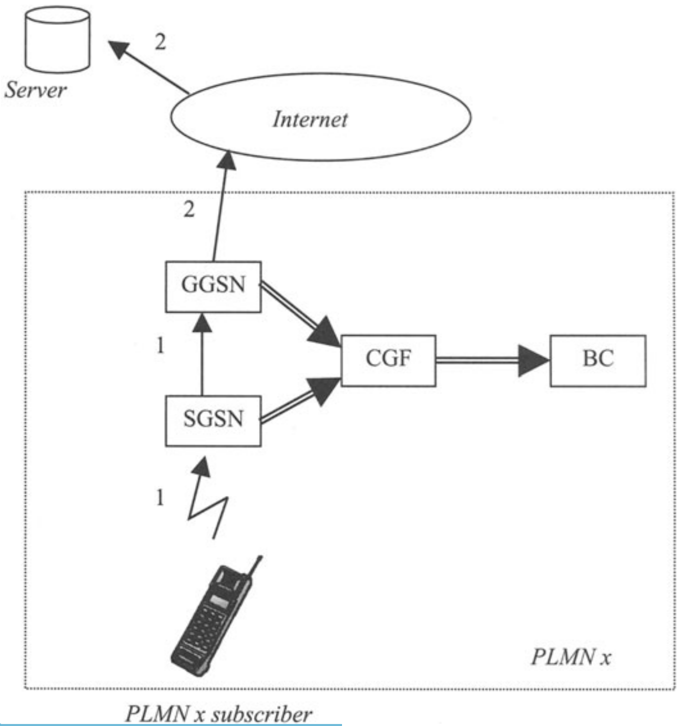


Figure 7.28 Access to an Internet server over the PS domain.

Table 7.2 Examples of call detail records.

SGSN record	GGSN record
IMSI of the mobile	IMSI of the mobile
IMEI of the terminal	Internet access point (APN)
Routing area identifier	Bytes transmitted on the uplink
Internet access point (APN)	Bytes transmitted on the downlink
Activation period of the PDP context	...
Bytes transmitted on the uplink	
Bytes transmitted on the downlink	
Bytes not transmitted (radio errors)	
...	

1 First of all, the mobile asks the SGSN to which it is connected to activate a PDP context. Depending on the Internet access point (APN: *Access Point Name*) requested by the mobile, it may also activate a context in the GGSN.
2 Once the contexts are activated, data can pass between the mobile and the information server. When the PDP context is deactivated (for example, at the end of the session on the server), the SGSN and the GGSN each send a call detail record to the BC through the CGF.

Table 7.2 shows a sample of the information contained in the records sent by the SGSN and the GGSN. Note that identification of the record is based on the IMSI, hence the importance of the security procedures used to ensure the confidentiality of the IMSI.

Incoming Call to a Roaming Mobile

In this type of call, illustrated in Figure 7.29, the mobile called is roaming and is therefore registered with a network y which is a VPLMN (*Visited PLMN*).

1 The call coming from the fixed subscriber is routed as far as the GMSC, the entry point of PLMN x.
2 The GMSC interrogates the HLR, which tells it that the subscriber to be reached is at present in PLMN y.
3 The GMSC creates a local call detail record, before routing the call to PLMN y.
4 On receiving the call, the GMSC will route it to the current MSC of the mobile. At the end of the call, the GMSC and the MSC will send a call detail record to the billing centre of PLMN y.

   The billing centre of PLMN y will then send a call detail record to the billing centre of PLMN x. This information will be used by the operator of PLMN x to pay part of the sum received back to operator y, as illustrated in Figure 7.30.

This repayment mechanism is quite complex, but it guarantees to each operator that all use of resources in the network will be associated with the collection of revenue. However, the mechanism contains a trap; when the mobile subscriber and the fixed subscriber are in the same country but the mobile subscriber is roaming

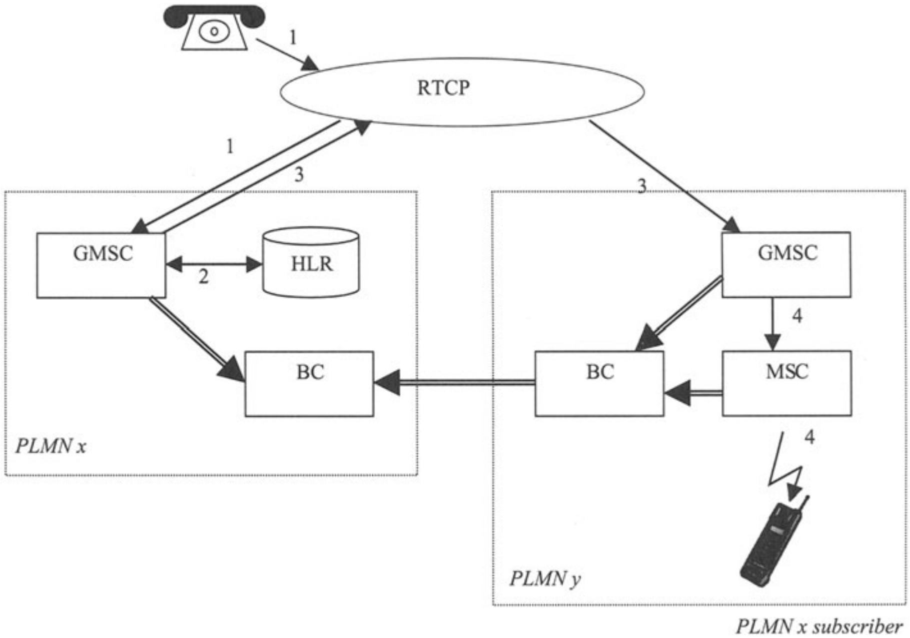


Figure 7.29 Call billing for a roaming subscriber.

away from his home country, the call will be billed at a high rate (because of the return journey between the two countries), while a simple local call would have sufficed. This situation is difficult to avoid, because it requires the implementation of optimised routing mechanisms.

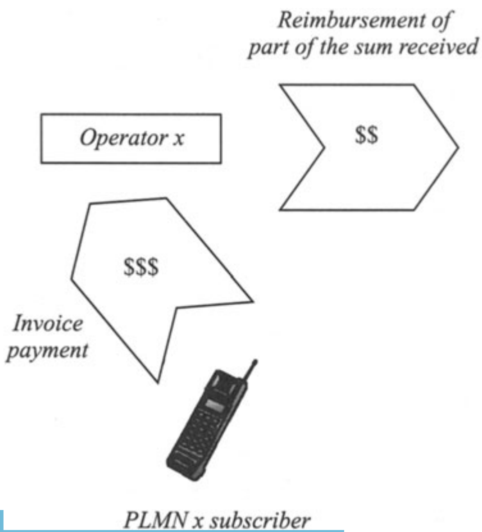


Figure 7.30 The reimbursement mechanism.

8 Mobility Management

8.1 Introduction

Mobility management is an essential function in cellular communication networks. In contrast to traditional fixed networks, the geographical position of the subscriber's terminal varies in the course of time. It is therefore necessary to include in a cellular communication network mobility management functions that allow a subscriber to be reached whatever his position in the network may be or, more generally, to ensure continuity of the service provided to the subscriber independently of his location in the operator's service area.

In a cellular environment, the size of the network cells and the speed with which users move around can be very varied. In the case of a motorway in open country, assuming a speed of 100 km/h and an average cell width of 20 km, the mobile will encounter a new cell every 12 minutes. On the other hand, if the motorway passes close to an urban centre with much smaller cells – 1 km, for example – the mobile will change cell every 36 seconds.

This period can be even smaller in the case of a pedestrian moving around in an urban environment with cells only a few hundred metres in diameter and subject to special propagation conditions (going round street corners, etc.).

It is therefore important to handle users' mobility efficiently, as much from the point of view of the user (any disturbance to the service provided to the user should be as small as possible) as from that of the network (the extra load generated by the mobility management functions must not disturb the running of the network).

From a simple viewpoint, mobility management can be separated into two parts according to the two possible states of a mobile that is switched on: idle mode and connected mode. The sections that follow describe the mechanisms used in the two modes. We will also address roaming (or inter-PLMN mobility) and intersystem mobility (such as the passage from a UMTS network to a GSM one).

8.2 Mobility in Idle Mode

In idle mode, the mobile is switched on but no connection is set up between the mobile and the network. A fortiori, no exchange of user data is possible in this mode.

This way of working does not exist in traditional wired telephony. In cellular telephony, even when the mobile is not engaged in a call, a certain level of activity must be maintained, as a consequence of the user's movement within the radio coverage.

The following subsections describe the principal operations employed in handling mobility in idle mode:

- geographical location of users in idle mode through the use of location zones;
- the cell selection and de-selection mechanisms used by a mobile in idle mode;
- broadcasting system information into the different cells of the access network.

8.2.1 Location Zones

In order to locate users whose mobile is in idle mode, the network is divided into geographical zones, called location zones, following the same principle as that used in GSM networks.

The standard allows for the possibility of managing mobility separately for packet-switched services and circuit-switched ones. We have seen, in fact, that the equipment in charge of mobility functions in the core network is distinct (VLR for the CS domain and SGSN for the PS domain); as a consequence, the standard defines different location zones for the two domains. The location zones for the CS domain are known as *Location Areas* (LA) and those of the PS domain are called *Routing Areas* (RA) (Figure 8.1).

Although they function independently, the CS and PS location zones obey certain rules:

- an LA can contain one or more RAs;
- an RA belongs to one and only one LA.

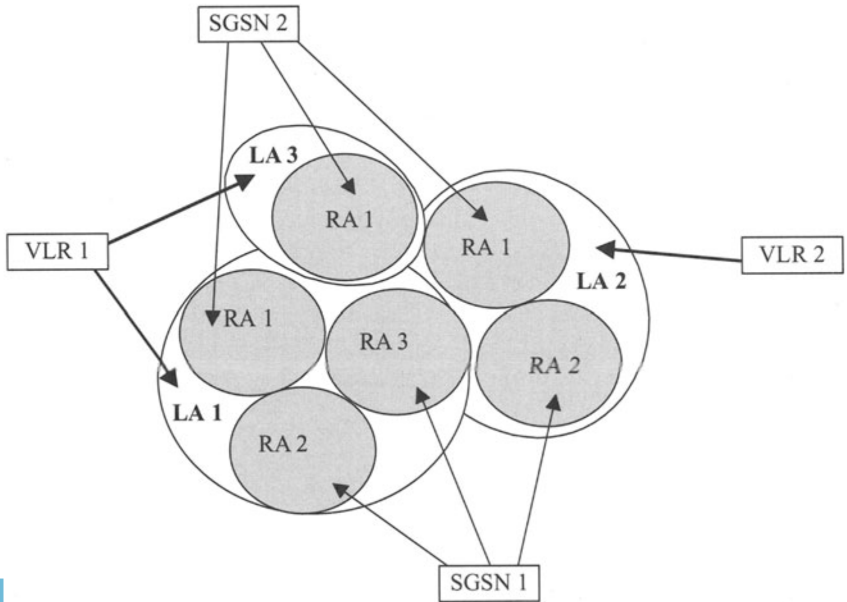


Figure 8.1 Examples of location areas and routing areas.

There is no a priori link between the location zones seen by the SGSN and the VLR. The standard requires only that an LA or RA is attached to only one VLR or SGSN. In contrast, the RAs forming part of the same LA can be managed by different SGSNs.

The terminal registers with the network when it is switched on, then updates the current zone as stored in the core network, every time it changes zone. The mobile knows the zone to which the current cell belongs thanks to the information broadcast on the beacon channel of each cell.

When the network wishes to set up a dialogue with one of the users, it uses a procedure called *paging*, which consists of broadcasting a message on the radio interface to alert the mobile. Splitting the cells of the network into location zones means that it is not necessary to send the paging message to all the cells in the network. It is sufficient to send the paging message to the location zone in which the mobile concerned is located.

The geographical extent and the positioning of the location zones are left to the unfettered judgement of the network operators. In general, they are determined by the geographical distribution of users over the cellular coverage. The size of the location zones is, in fact, a compromise between the signalling load imposed on the network and congestion on the paging channels.

The smaller the location zones, the greater the number of updates handled by the network, which can lead to a degradation of network capacity. On the other hand, the number of paging messages sent over the radio interface will be limited, since it depends on the number of cells in the location zones. If, however, the location zones are larger, the load on the paging channels will be that much greater.

Large location zones are therefore preferred in regions of low user density (the countryside or areas of low population density), small zones being used in urban areas characterised by a high density of users.

8.2.2 Updating Location Information

We have seen that the mobile must cause the network to update its record of the mobile's location zone every time that changes. For several reasons, this regular updating is not enough to ensure that the network runs faultlessly. The mobile is therefore also required to carry out the updating procedure from time to time even when the location zone has not changed. The frequency at which these updates are required is broadcast on the beacon channel of the cell and is fixed by the network operator. It can vary from 6 minutes to 25.5 hours.

When the mobile is switched off in the normal way, it must signal this to the network, so that the VLR's database can be updated. However, there is nothing to stop the user from simply removing the battery from the mobile. For the user, the result is the same: the mobile is off. For the network, however, the mobile is still considered as being in the idle state and located in the last location zone it signalled to the network.

Location Area Update for the CS Domain

Figure 8.2 shows an example of the procedure for updating the location area, in the case where the new and the old location areas are not handled by the same VLR.

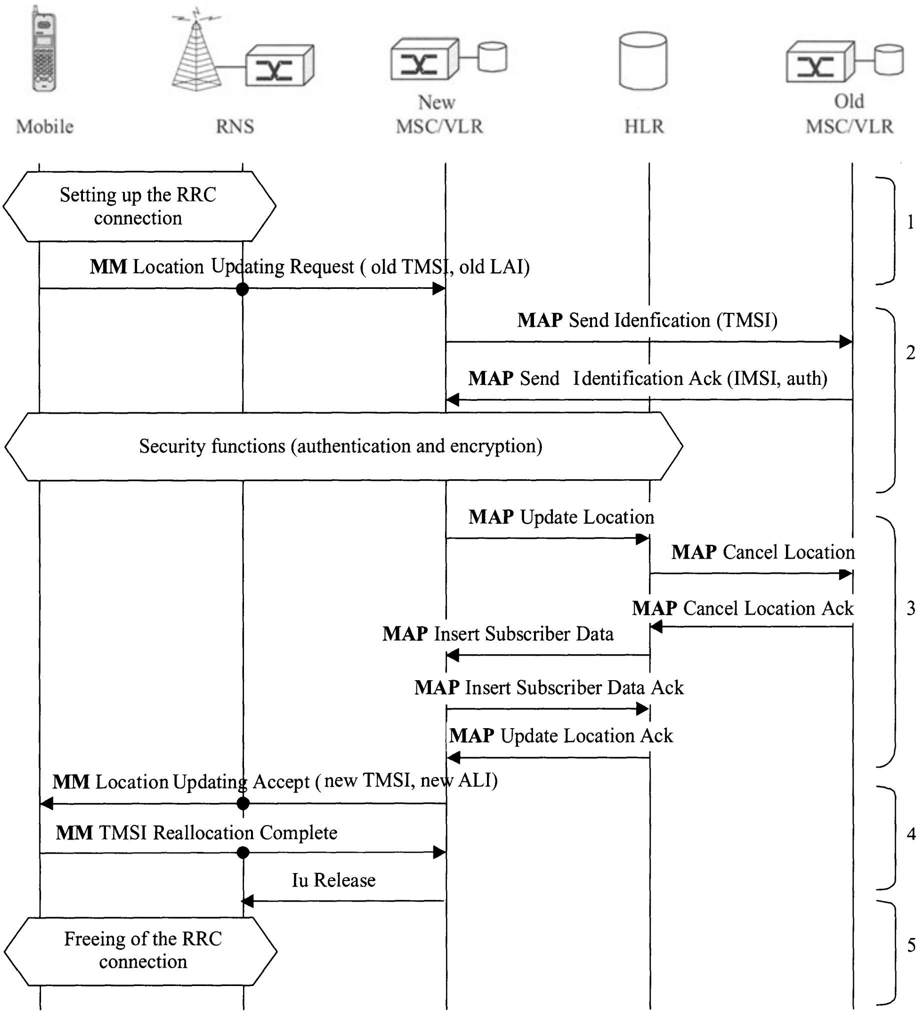


Figure 8.2 Example of location area updating.

1 Once the RRC connection between the mobile and the access network has been established, the request to update the location area is sent to the new VLR. The new VLR can determine the reference of the old VLR thanks to the fact that its identifier is provided by the mobile.

2 The old VLR is interrogated to obtain the real identity (IMSI) of the mobile and its authentication vector. The new MSC/VLR can then authorise the mobile to register and shift into encrypted mode. The encryption allows the information transmitted across the radio interface between the mobile and the network to be protected, especially the new TMSI, which will subsequently be allocated to the subscriber.

3 Once this information has been obtained, the new VLR informs the HLR of the change in the mobile's location. In its turn, the HLR will inform the old VLR, which will delete the database entry corresponding to the subscriber. During this phase, the HLR provides the new VLR with information about the services to which the subscriber has subscribed, by means of the *Insert Subscriber Data* procedure.
4 The mobile is informed that the procedure for updating the location area has been completed successfully. A new temporary identity (TMSI) is allocated to it by the new VLR.
5 The procedure having been completed, the MSC asks the access network to free the connection between the mobile and the network and de-allocate the resources used.

The messages exchanged between the different participants are prefixed by the name of the corresponding protocol layer:

- MM: these messages are transparent to the access network and are simply relayed by the RNS;
- MAP (*Mobile Application Part*): for the messages exchanged between nodes of the core network.

Routing Area Update for the PS Domain

Figure 8.3 shows an example of the updating of a routing area. The sequence of operations is almost identical to location area updating, even to the names of the messages; the procedures employed are functionally identical.

The protocols used are different, because the pieces of equipment and the data exchanged between them are different from the previous case. In updating the routing area, the following layers are used:

- GMM (*GPRS Mobility Management*), which is the equivalent of the MM layer for the PS domain;
- GTP (*GPRS Tunnelling Protocol*), which is the equivalent of the MAP for signalling exchanges between the nodes of the core network dedicated to GPRS.

8.2.3 Combined Updates

A combined update of the location area and the routing area is possible if the Gs interface between the SGSN and the MSC/VLR is supported by the core network or if the functions of the MSC/VLR are combined in a single element of the core network (we then speak of the UMSC, or UMTS MSC).

In this case, the mobile uses the routing area update procedure, just as if it were carrying out a simple routing area update. Since the routing area is a subset of the location area, updating the routing area means that the MSC/VLR (or the MSC/VLR part of the UMSC) automatically knows the mobile's location area.

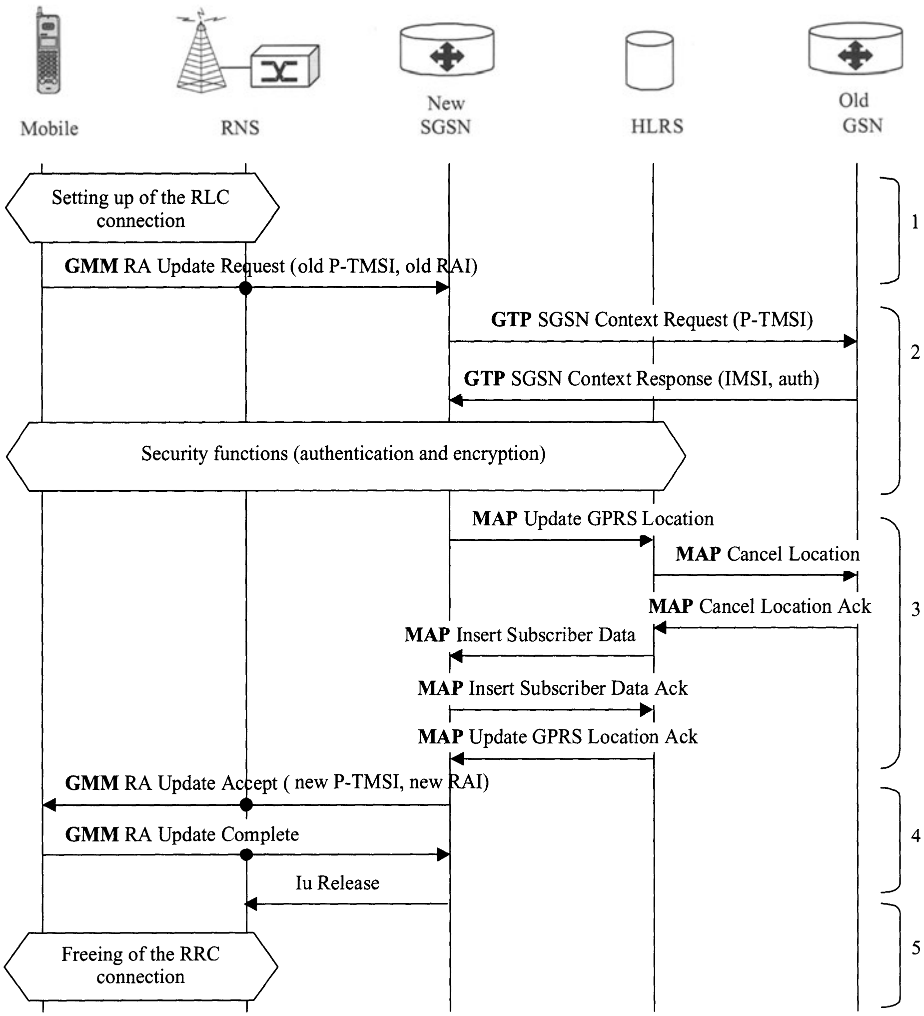


Figure 8.3 Example of a routing area update.

8.2.4 Cell Reselection

Cell reselection is the mechanism by which the mobile in idle mode is led to change the reference cell. The standard distinguishes two cases:

- the selection of the initial cell, carried out when the terminal is switched on;
- reselection of cells, carried out once the initial cell has been chosen.

In this subsection, we will describe only cell reselection. The initial selection process was described in Chapter 7, along with the other operations carried out when the mobile is switched on.

In idle mode, the mobile is attached to a particular cell of the network (the standard uses the phrase ‘to camp on a cell’). As a result of the user’s movements within the cellular coverage, the mobile may be led to select a new cell, if the current cell is no longer acceptable or if a new cell is seen by the mobile as being better than the current one. To do this, the mobile uses three mechanisms:

- the S (selection) or eligibility criterion. Only cells that satisfy the selection criterion are candidates for possible reselection. This is the same criterion as that used by the mobile when selecting the initial cell, as described in Chapter 7;
- the R (ranking) criterion, which allows the candidate cells to be ranked in order of desirability. The cell finally chosen by the mobile will be the highest ranking one;
- a measurement criterion, intended to limit the time the mobile spends carrying out measurements on the neighbouring cells.

Measurement Criterion

The energy consumption of mobiles has always been a worry for the manufacturers of mobiles and for network operators, all the more so in a complex radio environment, made up of cells using different frequencies or technologies, which requires the mobile to switch from one mode of listening to another.

In order to limit the time during which the mobile will be carrying out measurements on the neighbouring cells, the standard defines a measurement criterion based on the expression S_{qual}, which represents the quality of the current cell, as measured by the mobile. (The expression S_{qual} was defined in Section 7.1.4.)

Figure 8.4 shows how this criterion works.

If the value of S_{qual} is greater than a certain threshold, *Search*, broadcast by the network on the beacon channel, the mobile will not even try to measure the UTRAN cells neighbouring the currently selected cell. This will generally happen if the

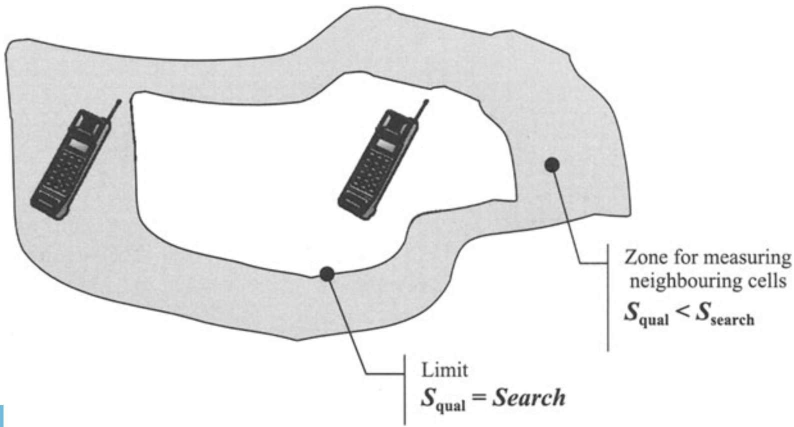


Figure 8.4 Measurement criteria in a UTRAN cell.

mobile is close to the antenna and/or the cell is lightly loaded. In this case, the mobile can be considered to be close to the centre of the cell and hence to have no reason for selecting another one.

In the contrary case, if the level of quality is below the *Search* threshold, the mobile will carry out measurements on the neighbouring cells, identify those that are eligible and rank them according to criterion R.

Criterion R

Criterion R is used by the mobile to rank the list of cells satisfying the eligibility criterion S. The cell that will be reselected is the one that ranks highest according to criterion R out of all those satisfying criterion S, for a minimum period whose value is specified by the network (*Treselection*). The mobile applies a different expression for criterion R according to whether it is a question of the current cell or neighbouring cells:

$$R_s = Q_{meas,s} - Qhyst_s$$

is the expression used for the current cell (*s* for serving).

$$R_n = Q_{meas,n} - Qoffset_{s,n} - TO_n$$

is the expression used for neighbouring cells (*n* for neighbouring).

$Q_{meas,s}$ and $Q_{meas,n}$ are the measures of quality of the current and neighbouring cells respectively. They are defined in the same way, depending on a parameter broadcast by the network on the BCCH, as either the ratio E_c / N_0 of the CPICH or the level of the signal received by the mobile on that channel.

Hysteresis, $Qhist_s$

In order to prevent the mobile reselecting cells too frequently, particularly at the boundary between two or more cells, a hysteresis threshold $Qhist_s$ is included in the expression for R_s. This threshold has the effect of favouring the current cell over the neighbouring ones.

Threshold, $Qoffset_{s,n}$

The threshold, $Qoffset_{s,n}$, applied to the R_n expressions, allows neighbouring cells to be favoured or penalised (depending on whether the value is positive or negative) independently of one another.

This mechanism can be used, for example, for a cell situated on the edge of a location area. Penalising a neighbouring cell situated in another location area restricts the number of update procedures carried out and hence the signalling load on the network.

Temporal Component, TO_n

A temporal component, TO_n, can be applied to the expression R_n, which serves to penalise the neighbouring cell for a specific period.

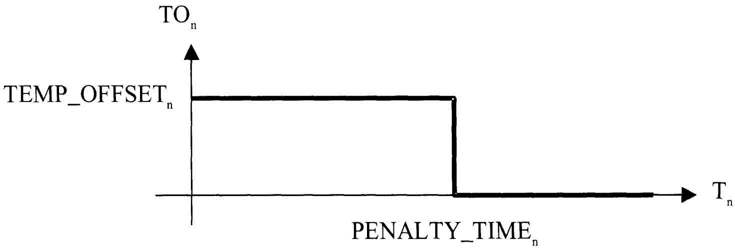


Figure 8.5 The TO_n component.

The temporal component is defined by

$$TO_n = TEMP\_OFFSET_n \times W(PENALTY\_TIME_n - T_n)$$

where $W(x) = 0$ for $x < 0$ and $W(x) = 1$ for $x \geq 0$.

To illustrate this more clearly, Figure 8.5 shows how the value of TO_n changes with time.

The period, T_n, for which the neighbouring cell n is penalised, is started by the mobile when the measurement $Q_{map,n}$ carried out on that cell exceeds ($Q_{map,s}$ + $Qoffset_{s,n}$).

The temporal component can be used in a microcellular environment to prevent a mobile in idle mode, moving rapidly, from selecting one of the cells of the microcellular layer.

The small cells belonging to the microcellular layer will only be seen as eligible by a mobile moving rapidly for a very short time. Thanks to the temporal component, TO_n, it is possible to penalise them in comparison with the larger cells (macrocells) whose R expression will not include a TO_n, component.

On the other hand, in the case of a mobile moving around slowly, i.e. one considered sufficiently stable to benefit from the microcellular coverage, the TO_n, component will eventually become zero. The corresponding cell will then be placed higher up in the list of eligible cells.

8.2.5 Role of the BCCH

We have already, in Chapter 5, considered the logical control channel BCCH (*Broadcast Control Channel*) and, more particularly, the associated physical channel, P-CCPH, as well as the broadcasting of system information into the cells of the UTRAN access network over the BCCH.

This subsection describes in more detail how the system information is structured before being broadcast on the radio interface and the use that is made of it in managing idle mode mobility.

Structure of the System Information

The UTRAN system information is structured hierarchically, as shown in Figure 8.6. The *System Information Blocks* (SIB) contain the system information broadcast

Figure 8.6 Hierarchical structure of system information.

by the network. The block structure allows information of a similar nature to be grouped together.

The *Master Information Block* (MIB) allows mobiles in idle mode to find out the types of information being broadcast (in fact, the SIB numbers) and their frequency. Thus, in contrast to GSM, the network can define different frequencies for different types of information, depending on the SIB number, and hence broadcast the more important system information more often.

The main types of SIB defined by the UMTS standard are the following.

- SIB 1 contains NAS level information including, for example, frequency of location updating.
- SIB 2 contains the identity of the URA associated with the cell.
- SIB 3 and 4 contain the parameters used by the mobile in cell selection and reselection.
- SIB 5 and 6 contain the configuration parameters of the common physical channels of the cell (for example, the parameters describing the PRACH).
- SIB 11 and 12 allow mobiles to configure the measurements to be carried out (i.e. the physical quantities to be measured and the lists of neighbouring GSM or UTRAN/FDD cells).
- SIB 13 contains information specific to ANSI-41 systems. This type of information shows that the UTRAN access network has been designed from the beginning to be connected also to a North American core network (the UMTS core network being based on the MAP standard).
- SIB 15 contains information used in finding the position of mobiles, for example, GPS information.

Broadcasting Information

We saw in Chapter 5 that the format of the P-CCPCH allowed 18 data bits per slot to be transported, which corresponds to fixed length data blocks containing 246 bits, transmitted over two radio frames (the type of channel coding and the size of the CRC are fixed by the standard) (Figure 8.7).

The information broadcast over the SIBs can be of very varied sizes. In order to make the best use of the bandwidth on the P-CCPCH, it is possible to segment or concatenate SIBs.

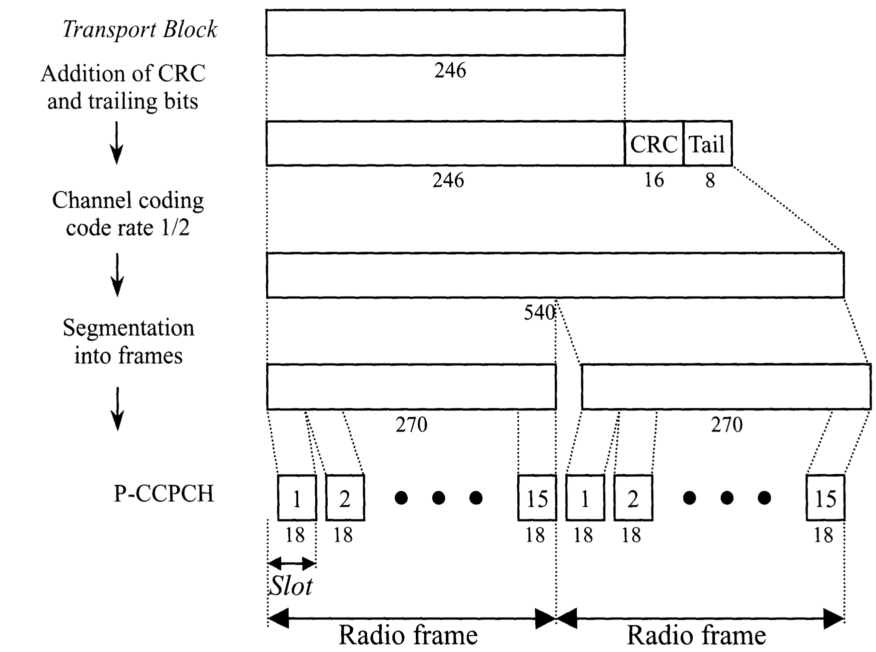


Figure 8.7 Physical layer operations on the BCH.

Figure 8.8 shows these different possibilities. Segmentation is carried out when the length of an SIB, for example SIB 3 in the figure, exceeds the transport block limit of 246 bits. In the case of small SIBs, such as SIB 1 and SIB 2 in the figure, it is possible to concatenate several of these into a single transport block.

Idle Mode Mobility Management Information

As a complement to the lists of neighbouring GSM and UTRAN cells, the different parameters used in the expressions for the R and S criteria used in the idle mode cell reselection algorithms are also broadcast on the BCCH of the current cell:

- $Qoffset_{s,\,n}$
- $Qhyst_{s}$
- *Qqualmin*
- *Qrxlevmin*

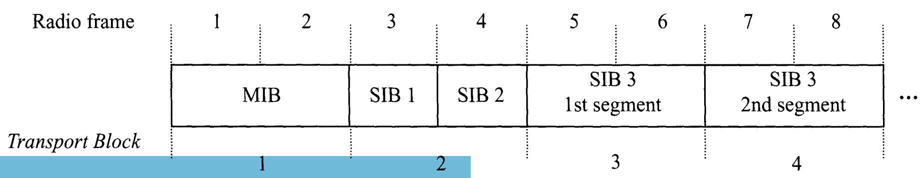


Figure 8.8 Example of SIB broadcasting.

- *$PENALTY\_TIME_n$*
- *$TEMP\_OFFSET_n$*
- *$Treselection_s$*

8.3 Mobility in Connected Mode

The purpose of the mobility functions in connected mode is to follow a user connected to the network in the course of his movements through the cellular coverage of the access network.

8.3.1 UMTS Principles

In connected mode, an RRC connection has been set up between the mobile and a UTRAN RNC, called the SRNC (*Serving RNC*). Thus, in contrast to the situation in idle mode, the mobile is referenced not only by the core network but also by the access network.

We saw, in Section 4.4.1, that there are several possible states of the RRC connection, corresponding to different levels of activity of the mobile. Depending on the state of the connection, the mobility of the user is placed under the control of either the mobile or the access network.

We will see later that, in the first case, the mechanisms employed are very close to those used for cell reselection in idle mode. In the case where mobility is under the control of the network, however, the term *handover* is used, which refers to the action of transferring the links between the mobile and the network from one piece of radio equipment to another (Table 8.1).

UTRAN Anchorage Point

We saw in Chapter 2 that mobility as seen by the core network was characterised by two main points.

In the case of CS services, mobility is characterised by the existence of an anchorage point in the core network, corresponding to the MSC/VLR that initiated the call. The reason for the existence of this anchorage point is connected with the billing mechanisms. The information necessary to bill the call is collected at the level of the anchor MSC. User data must therefore always pass through this MSC (Figure 8.9).

In the case of PS domain services, inheriting the principles of GPRS, mobility is characterised by the absence of an anchorage point in the core network. When

Table 8.1 RRC states and mobility control.

RRC state	Mobility control	Procedure used
CELL_DCH	Network	Handover
CELL_FACH	Network or mobile	Handover or cell update
CELL_PCH	Mobile	Cell update
URA_PCH	Mobile	URA update

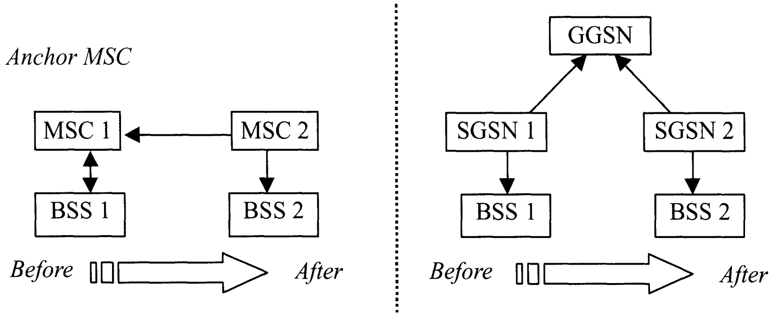


Figure 8.9 Mobility and transfer of user data.

subscribers are led to change SGSN (PS domain equivalent of the MSC/VLR), the information associated with them, such as the attributes of the PDP contexts set up, is transferred from the old SGSN to the new one. For this reason, user mobility in GPRS leads to procedures that are slow and costly in terms of signalling exchanges in the network. The GPRS billing mechanisms necessitate the collection of information situated in different nodes of the network, the SGSN and the GGSN.

For mobiles in connected mode, the UMTS standard offers the possibility of using the RNC supporting the RRC connection set up with the mobile as the point of anchorage for mobility in the UTRAN access network. Having a point of anchorage in the UTRAN allows the signalling load in the network to be reduced by shielding it from part of the inter-RNC mobility.

It is therefore theoretically possible to keep the same SRNC throughout the duration of the call, whatever the position and movements of the mobile in the network. Of course, this is not desirable, because of the transmission costs in the access network, especially at the level of the Iur interface.

In order to optimise routing, the UTRAN can change the point of anchorage of the RRC connection, using a procedure known *relocation*, described in Section 8.3.4 below (Figure 8.10).

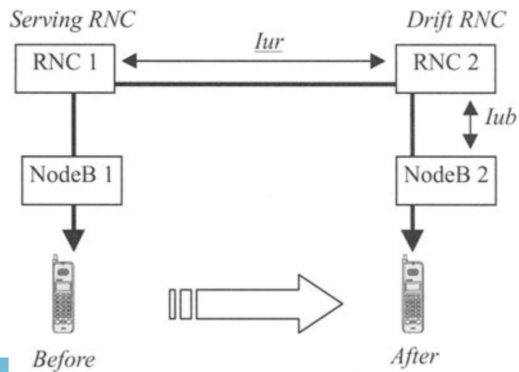


Figure 8.10 Serving and drift RNCs.

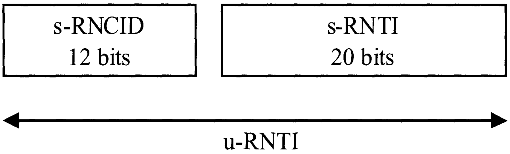


Figure 8.11 Structure of the u-RNTI.

The possibility of using the SRNC anchor point requires several supplementary definitions. The RNC controlling the new cell to which the mobile is attached is called the *drift* RNC (DRNC). The role of the DRNC is to handle the routing of user data and signalling information between the new Node B and the SRNC.

The DRNC also plays a role in the allocation of resources in the new cell belonging to Node B2. Since the SRNC does not control the cells of Node B2, it cannot directly allocate resources in Node B2 for the mobile concerned. This function is therefore carried out by the DRNC, at the request of the SRNC.

A new temporary identity for referencing mobiles at access network level, the u-RNTI (*UTRAN Radio Network Temporary Identity*) is in some ways an access network equivalent of the TMSI. It is made up of two fields (Figure 8.11):

- the s-RNCID (*Serving RNC Identity*), which uniquely identifies the SRNC within the access network;
- the s-RNTI (*Serving Radio Network Temporary Identity*), which uniquely identifies the mobile at the level of the SRNC.

The u-RNTI remains generally unchanged throughout the life of the RRC connection, except in the case of relocation.

8.3.2 Control of Mobility by the Terminal

In three of the states of the RRC connection (CELL_FACH, CELL_PCH and URA_PCH), the mobility of the user is controlled by the mobile, in a way identical to that used in idle mode.

The states CELL_PCH and URA_PCH are in some sense dormant states, in that no data is exchanged between the mobile and the network, despite the existence of a connection. In the CELL_FACH state, user data can be transferred between the mobile and the network. Depending on the volume of data transmitted, the network can ask the mobile itself to control the cell changes, as it would do in the CELL_PCH and URA_PCH states. As in idle mode, the mobile must inform the network, especially the SRNC, of changes in position within the cell coverage. In the CELL_FACH and CELL_PCH states, the mobile signals each cell reselection to the network, using the *cell update* procedure. In the URA_PCH state, the mobile must signal every change of URA to the SRNC, using the *URA update* procedure.

The idea of the URA (*UTRAN Registration Area*) was described during the description of the RRC layer. The URA is, in fact, a grouping of cells. The concept of the URA brings flexibility to the functioning of the network. Mobiles identified by the network as not very active or as moving very quickly can be placed in the

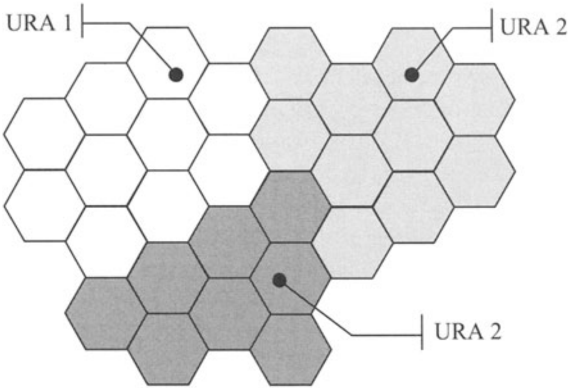


Figure 8.12 URAs and cells.

URA_PCH connection state, in such a way as to reduce the network load arising from procedures of the *cell update* type (Figure 8.12).

As in the case of updates to the LA and RA, the network can ask the mobile to update the cell or URA regularly, even when it has not changed. The frequency of such updates is broadcast by the network on the BCCH of every cell; it can vary from every 5 minutes to every 12 hours.

Figure 8.13 shows an example of the procedure for updating a cell. In the CELL_PCH state, the mobile cannot send messages to the network. A temporary transition to the CELL_FACH state is therefore carried out during the cell update procedure.

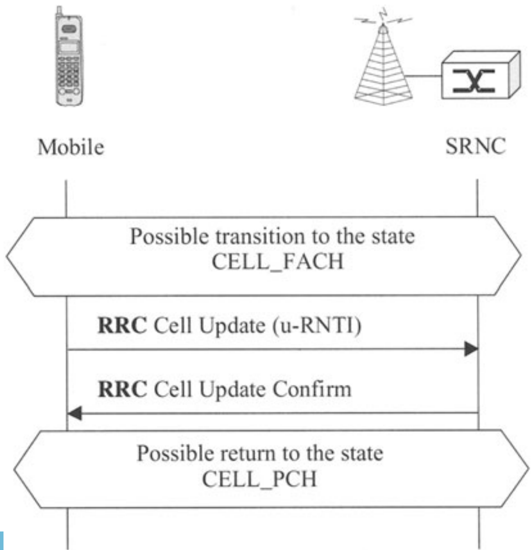


Figure 8.13 The cell update procedure.

The procedure for updating the URA is very similar. While the procedure is going on, the mobile must again move into the CELL_FACH state.

Using the SRNC as a point of anchorage for mobility in connected mode allows the mobility procedures of the core network and the access network to be completely separated. In this way, the core network is totally independent of the update procedures of the access network (*cell update* and *URA update*). From the point of view of the core network, the point of access to the mobile is the SRNC. This access point remains unchanged so long as the SRNC does not initiate any relocation.

The cell reselection criteria applied by the mobile are based on the same algorithms as those used in idle mode (i.e. the R and S criteria described in the previous sections), but use a set of parameters that can be different from that used in idle mode. (The reselection parameters for connected mode are broadcast on the BCCH of the current channel using different messages.)

8.3.3 Mobility Control by the Network (Handover)

Handover in GSM

In GSM technology, the mobile is in communication with only one piece of radio equipment at a time. A cell change therefore translates into a switch from the old channel to the new channel allocated by the new cell in which the mobile is now located.

This way of switching the channel in use results in cutting communication for the period during which the old channel is not in use and the new one is not yet usable. This gap, of the order of 100 ms is sufficiently small not to disturb a telephone conversation. However, in the case of data communication, it can happen that one or more blocks are lost. Thanks to the RLP (*Radio Link Protocol*) set up between the mobile and the IWF (*InterWorking Function* or point of access to the telephone network) and applied to data services in non-transparent mode, the lost data is re-transmitted (Figure 8.14).

Because of this brief interruption in the transmission of user data, handover in GSM networks is known as *hard handover*, in contrast to the handover techniques used in CDMA networks.

Handover in CDMA

CDMA networks use a different handover technique, called *soft handover*, because transmission is not interrupted when the user changes cell.

The soft handover technique allows the mobile to be connected simultaneously to several base stations. The same user information is transmitted simultaneously on the radio links used, which allows both the mobile and the network to benefit from an increase in diversity. In the upward direction, the signal transmitted by the mobile, using the scrambling code C_s is received by different Node Bs for subsequent onward transmission to the SRNC. In the downward direction, the Node Bs involved in the soft handover transmit simultaneously the same user information to the mobile. Each Node B uses the scrambling code belonging to its cell (Figure 8.15).

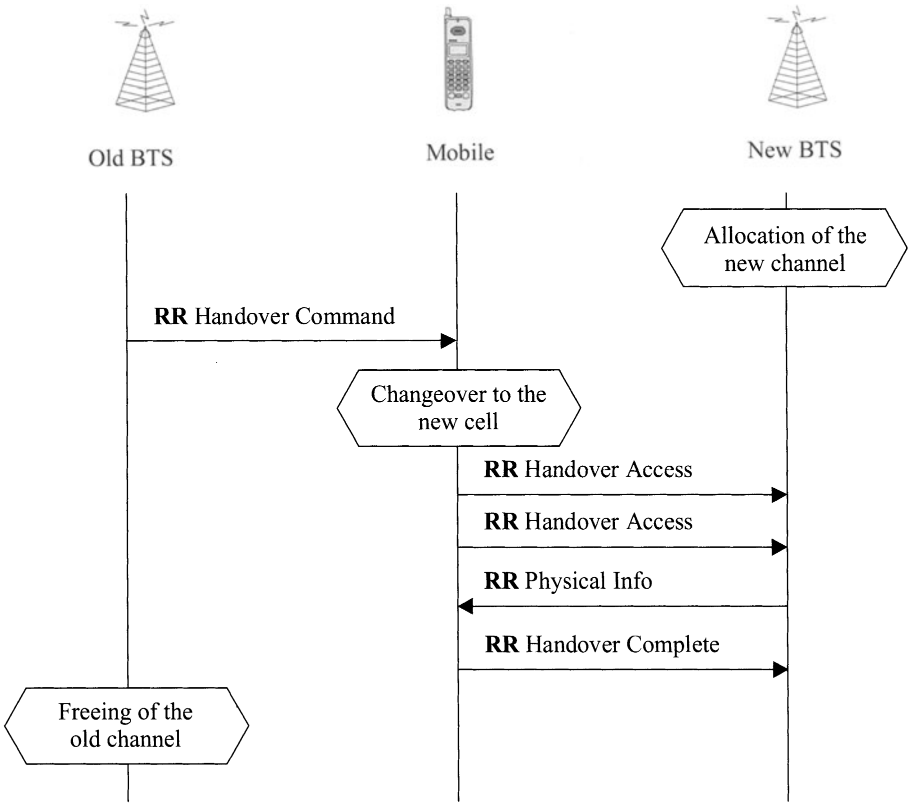


Figure 8.14 Example of inter-BTS handover in GSM.

The set of radio links in use simultaneously between the network and the mobile is called the *active set*. The active set can contain up to six radio links.

It is possible that the cells involved in the soft handover belong to the same Node B. In this case, the soft handover links are supported by the sectors of the Node B. This is sometimes called *softer handover* (Figure 8.16).

When the mobile begins to move away from its current cell, and therefore gets closer to one or more neighbouring cells, one or more links are set up with the neighbouring cells considered the most favourable. If the mobile continues to move away from its current cell, the radio link for that cell will eventually be shut down. The mobile then moves into a different cell without the communication channel between the mobile and the network being cut.

Figure 8.17 shows an example of the algorithm used by the SRNC to manage the active set. The mobile measures the ratio E_c / N_o of the CPICH of each neighbouring cell. The SRNC decides to add a cell to the active set when this ratio passes a threshold S1 for a period of at least T_{add} seconds. Conversely, the cell is taken out of the active cell when the ratio E_c / N_o falls below the threshold S2 for at least T_{sup} seconds.

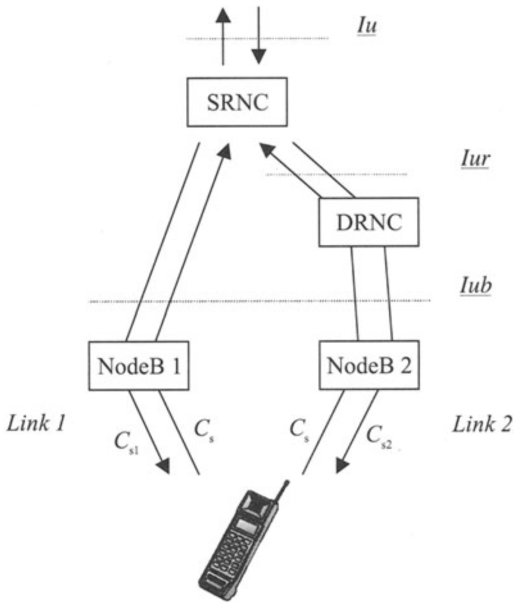


Figure 8.15 Soft handover involving two Node Bs.

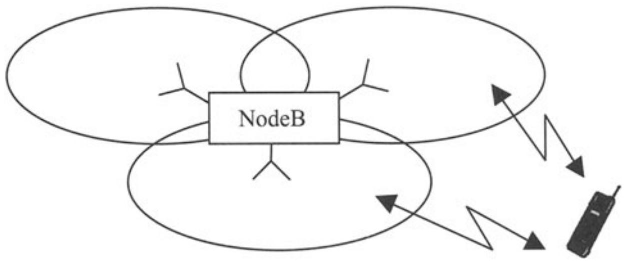


Figure 8.16 Softer handover.

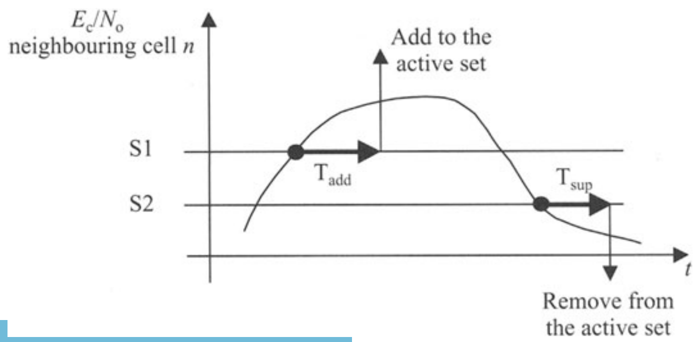


Figure 8.17 Management of the active set.

Figure 8.18 shows an example of the soft handover procedure corresponding to Figure 8.15. Before the procedure is started, only one radio link exists between the mobile and the network. The diagram shows the addition of a second link, via a drift RNC controlling the Node B2.

1 On the basis of the measurements sent back by the mobile, the SRNC decides to add a link to the active set between the mobile and the Node B2.
2 The SRNC instructs the Node B2, through the DRNC controlling it, to set up a new link with the module. The corresponding virtual circuits are also set up on the Iub and Iur interfaces (*Establish Request* and *Establish Confirm* in the AAL2 signalling set).
3 Once the radio link is correctly configured and the transmission route on the terrestrial network has been set up, the SRNC informs the module of the new link in the active set.

This technique is not just an improvement on the mechanisms of handover in GSM; it is indispensable in CDMA networks, in order to limit interference due to mobiles situated on the border of a cell. In a TDMA/FDMA network, this interference is naturally limited by the frequency differences between neighbouring cells. In a

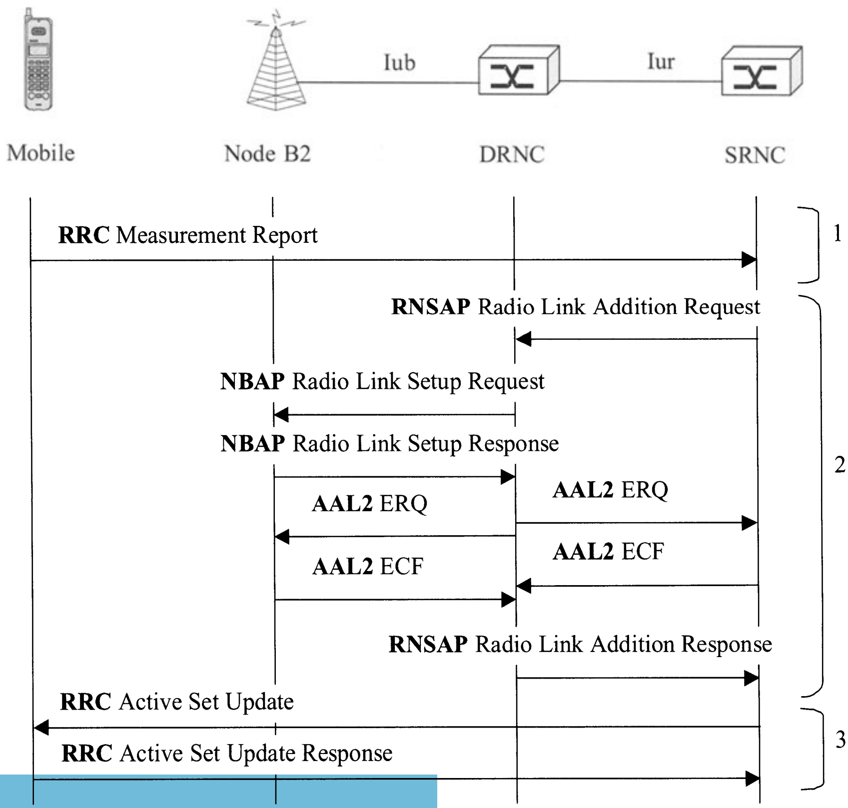
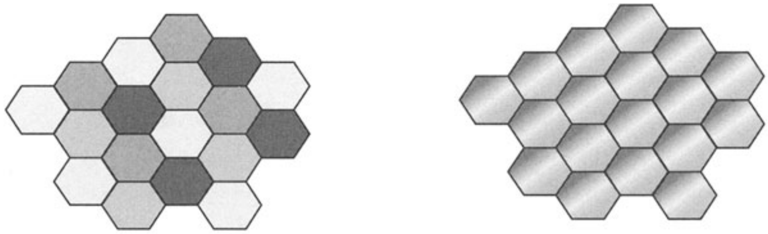


Figure 8.18 Adding a soft handover radio link.



Reuse pattern in GSM

In CDMA, the same frequency is reused

Figure 8.19 Use of frequencies in GSM and UMTS.

CDMA network, all cells use the same frequency. A handover mechanism similar to that of GSM would cause mobiles situated on cell borders to use transmission power that was too high. The level of interference would increase, leading to a reduction in network capacity (Figure 8.19).

The mechanism used in CDMA allows the transmission power used by mobiles in a soft handover situation to be reduced. When the signal transmitted by the mobile is received by several base stations, the transmission power of the mobile can be adjusted by the network, depending on the signal strength received by the base station that is receiving the best signal. The transmission power is thus lower than if the mobile was in communication with only one base station.

8.3.4 Relocation

The relocation procedure is used to move the UTRAN anchorage point of the RRC connection set up between the mobile and the network. In the UMTS specifications, this procedure carries the name *SRNC relocation* because the effect is to change the current SRNC of the mobile. It is used in the following two cases:

- optimisation of information routing in the UTRAN;
- support of hard handover.

Routing Optimisation

We have seen that the soft handover mechanism of the UTRAN allows the mobile to move freely in the network, while still remaining connected to the same SRNC. It is not, however, desirable to retain the same SRNC during the whole length of the RRC connection for all types of communication. Such a choice would lead to high transmission costs on the terrestrial interfaces of the UTRAN.

Figure 8.20 shows the relevance of relocation in optimising user transmissions between different RNCs in the UTRAN. When no relocation is carried out, it can happen that a large number of connections between mobiles and SRNCs require the use of a DRNC. In the left-hand part of Figure 8.20, the four mobiles shown are in this situation, with RNC1 and RNC2 playing the role of SRNC or DRNC

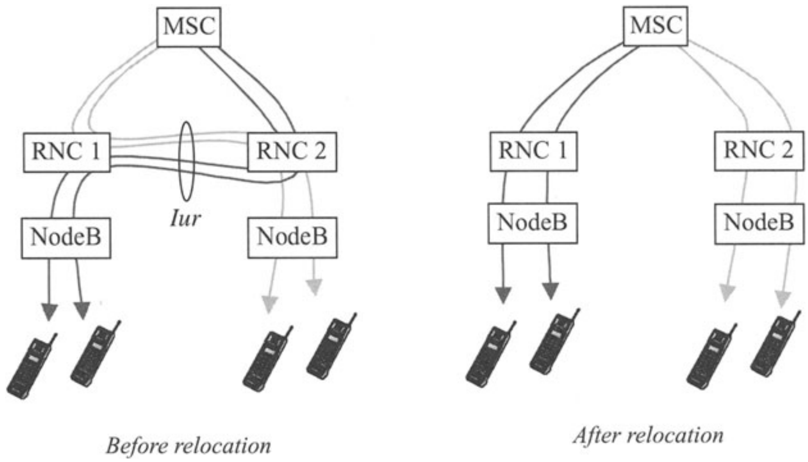


Figure 8.20 Optimisation and relocation.

according to circumstances. Relocation allows the SRNC used for each connection to be redefined and the load on the Iur interface can thus be optimised.

In Section 4.2, we saw that there are certain states of the RRC connection in which little or no user data is exchanged (in particular, the states CELL_FACH, CELL_PCH and URA_PCH). Because of the significant delay that can occur between the moment at which the connection between the mobile and the SRNC is created, and the start of the transfer of user data between the mobile and the network, the SRNC can be a very long way from the Node B, which incurs a high transmission cost. In this case the relevance of relocation is still more apparent.

The algorithm by which the SRNC decides to undertake data routing optimisation is not specified in the UMTS standard. The standard states only that the SRNC relocation procedure should not be initiated if there exists at least one radio link between the mobile and one of the Node Bs controlled by the SRNC.

Intra-MSC relocation

Figure 8.21 shows the relocation procedure. In this example, only one MSC is in use (the current RNC and the target RNC are connected to the same MSC). The arrows show the route taken by the user data. Before relocation, user data travels from the SRNC to the mobile via the DRNC. After relocation, the DRNC has become the new SRNC.

The relocation procedure is made up of three main parts (Figure 8.22).

1 *Preparation*. The current SRNC has decided to start the relocation procedure. A request is sent to the MSC, which asks the target SRNC to reserve the necessary resources. In this case, the only resources necessary are new AAL2 channels between the target SRNC and the MSC. During this phase, data regarding the configuration of the radio protocol layers (PDCP, RLC, MAC) is sent from the current SRNC to the target. The MSC is informed of the completion of the preparation phase by the message *relocation request ack*.

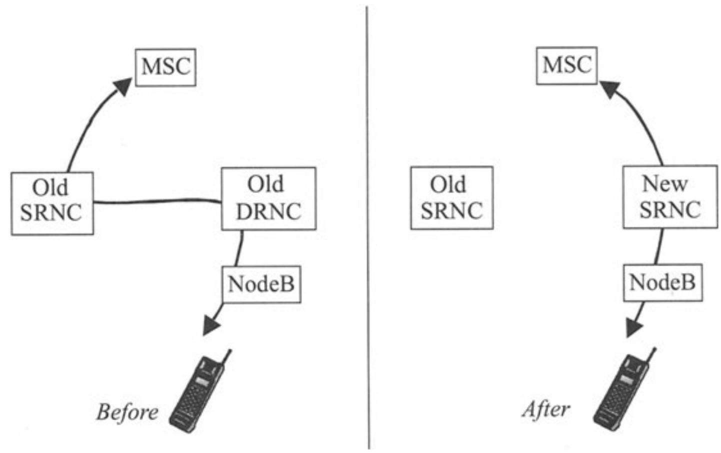


Figure 8.21 Intra-MSC relocation.

2 *Execution*. The message *relocation commit* signals to the RNC that it can consider itself as the new SRNC. The new SRNC can then allocate a new u-RTNI (temporary reference for the mobile in the SRNC). The MSC is informed of the completion of this phase by the message *relocation complete*.
3 *Resource de-allocation*. Since the relocation is now complete, the MSC requests the old SRNC to release the old resources no longer required. In this case, only the AAL2 channels need to be released.

As Figure 8.22 shows, routing optimisation in the fixed part of the UTRAN network has no effect on the transmission links used over the radio interface. This type of relocation is in effect transparent (up to the reallocation of the RNTI) to the Node Bs supporting the radio links and to the mobile and, importantly, to its user.

Relocating a connection supporting a packet-switched service is almost identical. The only differences are:

- replacement of the MSC/VLR by the SGSN;
- absence of AAL2 signalling on the Iu interface.

Inter-SGSN relocation

When the new and the old SRNCs are attached to different SGSNs, the relocation procedure requires signalling exchanges between the new and the old SGSN, using the GPRS tunnelling protocol (GTP). The GGSN, the point of access to the Internet, remains unchanged during the procedure (Figure 8.23).

Inter-SGSN relocation involves three phases (Figure 8.24).

1 *Preparation*. The request for relocation issued by the current SRNC is transmitted to the target SGSN by means of the *Forward Relocation Request*. A positive

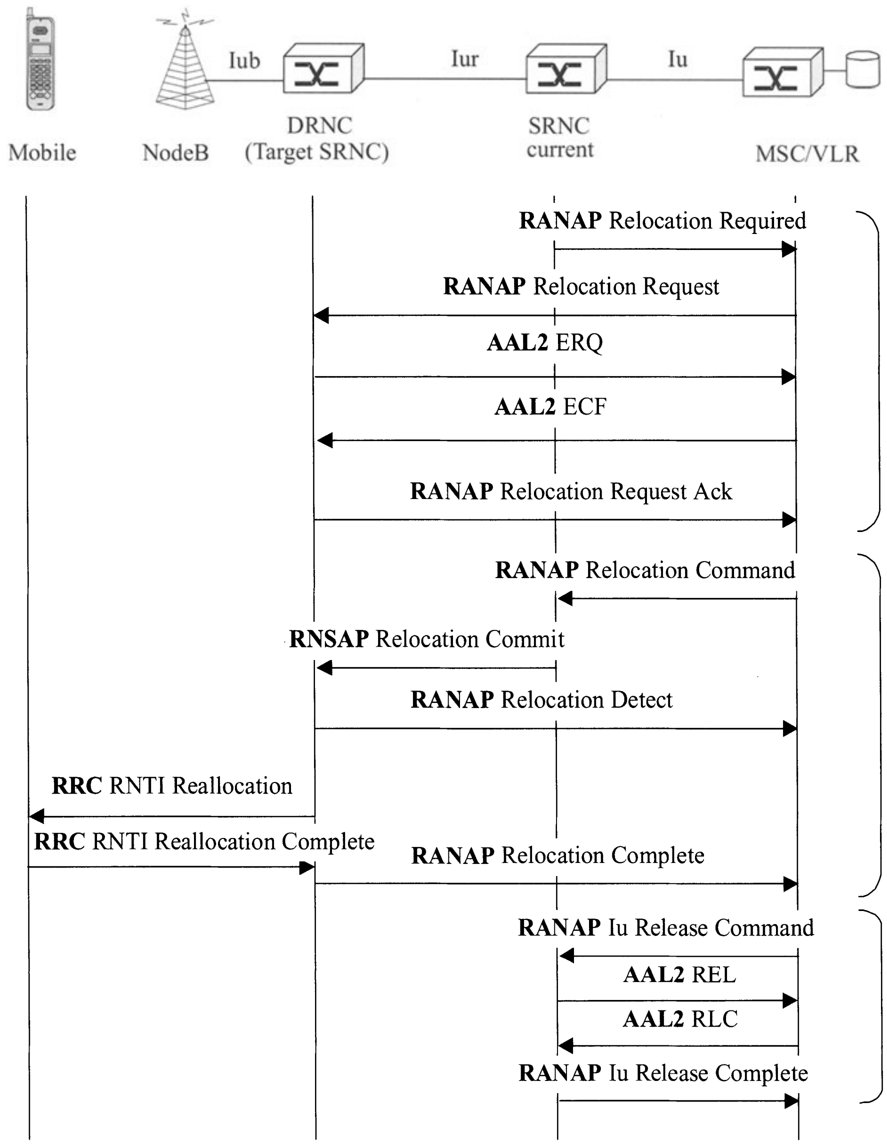


Figure 8.22 The relocation procedure.

response is sent by the target SGSN when the target RNC has allocated the new resources required.

2 *Execution*. During this phase, the old SGSN is informed of the new SGSN to which the mobile is attached. The GGSN must then transfer the user data coming from the external network to the new SGSN.

3 *Resource de-allocation*. This phase is identical to that of the preceding case.

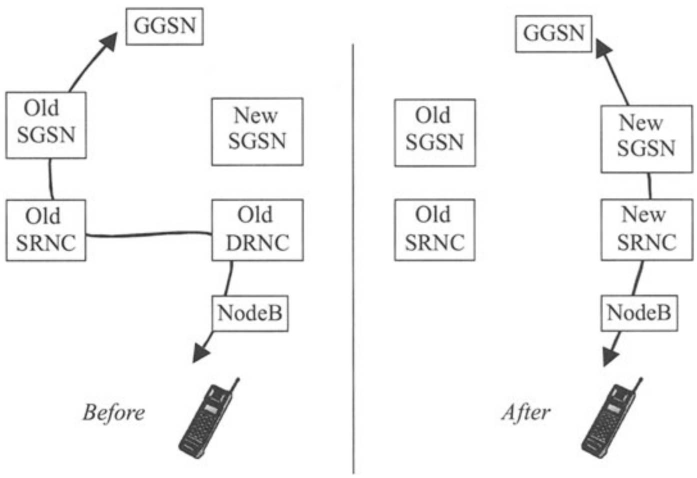


Figure 8.23 Inter-SGSN relocation.

Relocation Simultaneously Involving PS and CS Domains

When the mobile is simultaneously using both PS and CS services, two connections are set up between the SRNC and the nodes of the core network, one between the SRNC and the MSC, and one between the SRNC and the SGSN. Two relocation procedures are then carried out by the SRNC, the first with respect to the MSC and the second with respect to the SGSN. This double relocation requires special treatment on the part of the current SRNC and the target SRNC; if one of the two connections, CS or PS, cannot be relocated, the whole procedure must be abandoned.

Support for Hard Handover

We have seen that soft handover provides mobile users with continuous transmission over the radio interface. It may happen, however, that the network has recourse to different handover procedures, in one of the following cases, for example:

- The Iur interface is not available. In this case, soft handover can only take place between cells belonging to the same RNC. If the mobile strays too far from the cells controlled by it SRNC, it becomes necessary to use a hard handover.
- Inter-frequency handover. Soft handover cannot be carried out between cells using different frequencies. Depending on the structure of the network and the traffic partitioning strategy used by the operator, it may be necessary to use hard handover in this case.
- FDD/TDD handover. As in the previous case, the mobile can only use one access technology at a time.

If the current and target cells do not belong to the same RNC, it is necessary to carry out an SRNC relocation procedure at the same time as the handover. In fact,

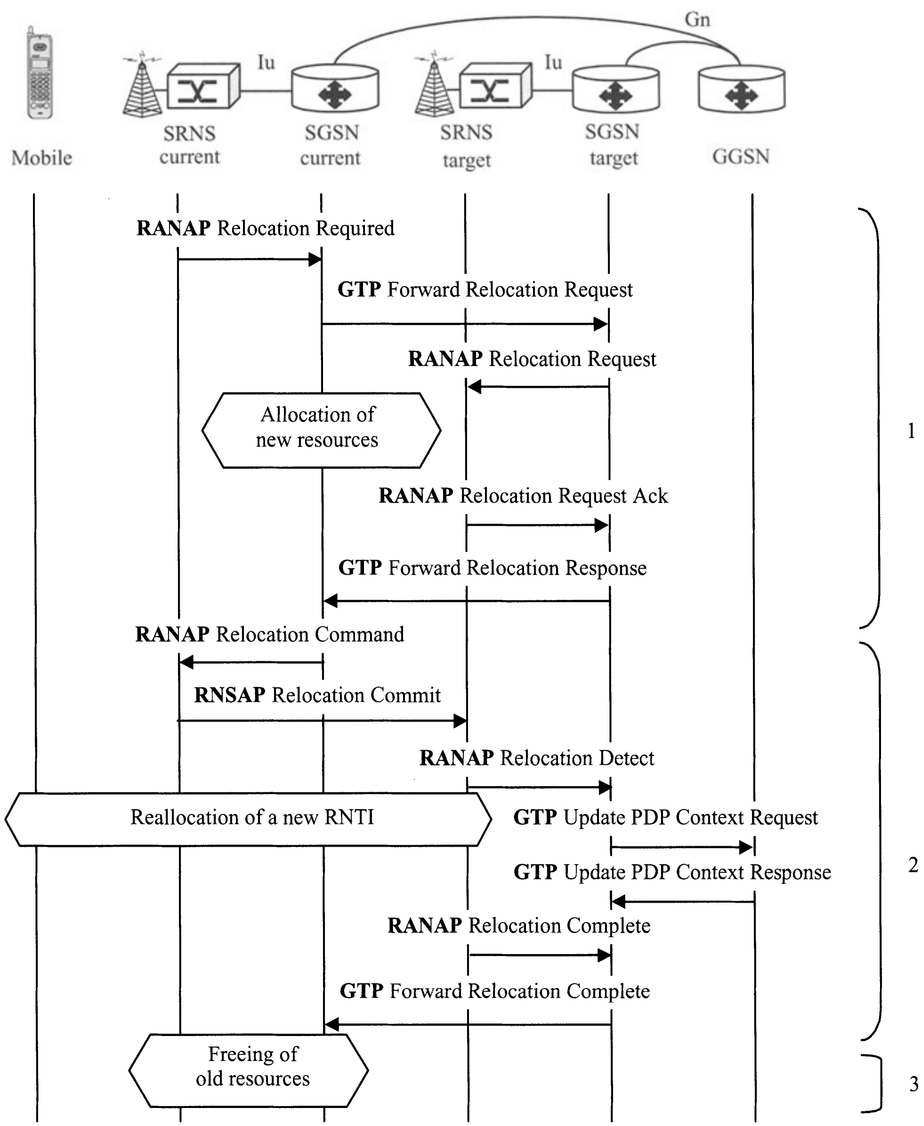


Figure 8.24 Inter-SGSN relocation.

this is done using a mechanism identical to that used in GSM in the case of external handovers (or inter-BSC handovers) (Figure 8.25).

In contrast to the case where relocation is carried out in order to optimise routing, in this case the mobile is involved in the processes, because new radio resources in the target cell are allocated to it.

The example in Figure 8.26 describes the signalling exchanges between nodes of the UMTS network in the case of simultaneous relocation and hard handover.

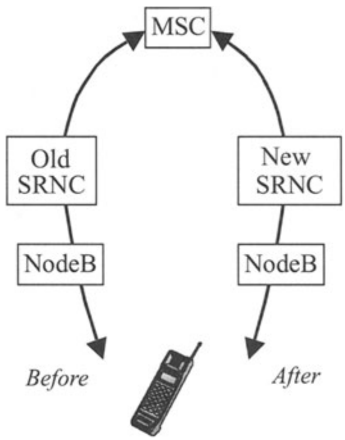


Figure 8.25 Relocation and hard handover.

The process involves the same phases as the earlier relocations:

1 *Preparation*. In addition to the resources allocated on the Iu interface, the target RNC must also allocate an AAL2 virtual circuit on the Iub interface with the target Node B. Further, a new radio link is activated in the target cell, by means of the *radio link setup* procedure of the NBAP protocol.
2 *Execution*. During this phase, the SRNC must instruct the mobile to change cell. The execution phase is complete when the mobile has successfully locked on to the new cell and a new u-RNTI has been allocated.
3 *Resource de-allocation*. The old radio link and the AAL2 virtual circuit are released by the old SRNC.

8.4 Intersystem Mobility

From the start of UMTS deployment, mobility management between the UMTS and GSM segments of cellular networks will be a priority, because of the lack of national UMTS cover and the need to provide service continuity to the network users. Cellular networks will need to be careful to guarantee user mobility from one access technology to another, not only in idle mode but also during a call.

Support for intersystem mobility imposes a number of special requirements on the UMTS network, which have been included in the specifications since the first version of the 3GPP standard. In the rest of this section, we will look at the developments necessary in the standard and in GSM networks in order to support intersystem mobility.

8.4.1 Neighbourhood Management

Whether we are considering reselection in idle mode or handover in connected mode, the decision to do it depends on the measurements carried out by the mobile on the cells neighbouring the current one.

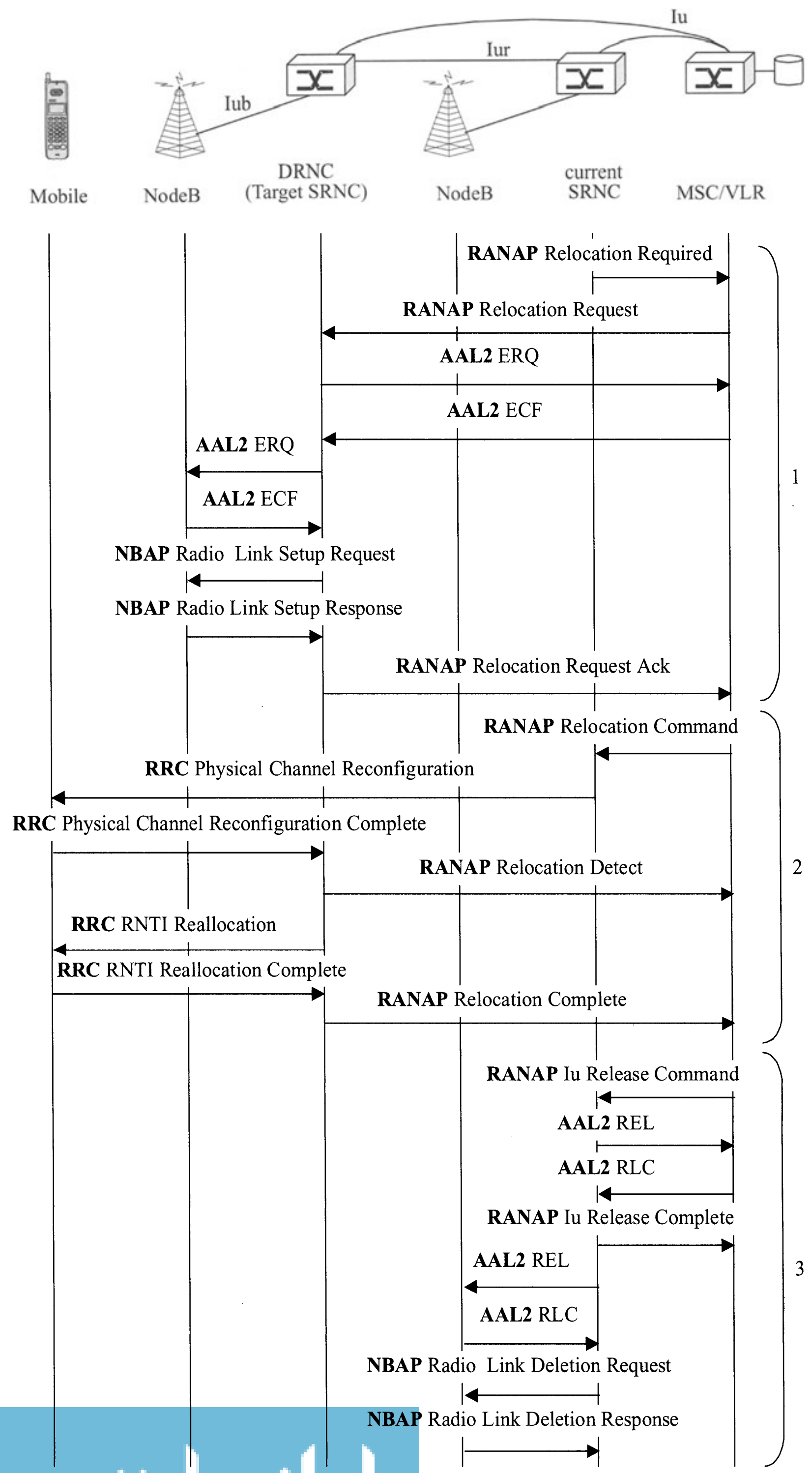


Figure 8.26 Relocation and hard handover.

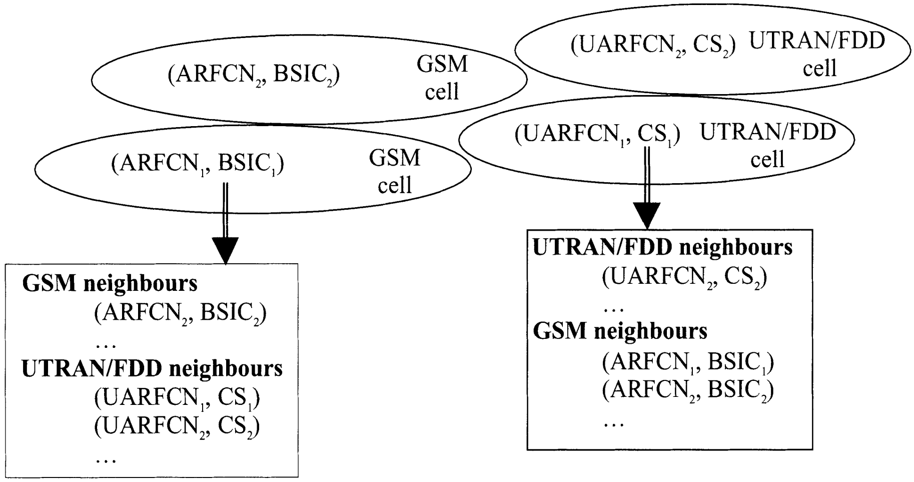


Figure 8.27 Broadcasting neighbourhood data in a multimode network.

In a multimode network, therefore, the network must not only provide the mobile with a list of neighbouring cells operating in the mobile's current mode, but also with a list of cells using the other access technology (Figure 8.27).

In each GSM cell, as well as the list of GSM cells, the network broadcasts on the broadcast control channel the information that idle mode mobiles need to carry out measurements on the neighbouring UTRAN/FDD cells. For each UTRAN/FDD neighbour, the network tells the mobile the cell's carrier frequency (UARCFN: *UTRAN Absolute Radio Frequency Channel Number*) and the primary scrambling code used on the P-CCPCH (the physical channel used for broadcasting). Similarly, in each UTRAN/FDD cell, the network must broadcast the (ARFCN, BSIC) pair corresponding to each neighbouring GSM cell. The ARFCN is the GSM equivalent of the UARCFN. The BSIC (*Base Transceiver Identity Code*) serves to identify the GSM cell.

8.4.2 Intersystem Reselection

Before measuring the level of the BCCH of a neighbouring GSM cell, the UMTS mobile in idle mode must first decode two sub-channels of the GSM BCCH:

- the FCCH (*Frequency Correction Channel*), used for frequency adjustment of the mobile;
- the SCH (*Synchronisation Channel*), which contains the BSIC information associated with the cell and which thus allows the mobile to satisfy itself about the identity of the cell it is measuring.

Once the level measurement has been carried out, the mobile uses the same reselection criteria as those described in Sections 7.1.4 and 8.2.4 (i.e. criteria S for eligibility and R for establishing the order of preference), slightly adapted to take account of the different nature of the physical quantities measured for a GSM cell.

Thus, for a GSM cell the S criterion takes the form ($S_{\mathrm{rxlev}} > 0$), S_{qual} not being taken into account for a GSM cell. Srxlev is defined by

$$S_{\mathrm{rxlev}} = Q_{\mathrm{rxlevmeas}} - Q_{\mathrm{rxlevmin}} - P_{\mathrm{compensation}}$$

In this case, $Q_{\mathrm{rxle2vmeas}}$ represents the level of the signal received on the GSM cell's beacon channel, BCCH.

Similarly, the ordering criterion R for a GSM cell is defined by

$$R_n = Q_{meas,n} - Qoffset_{s,n} - TO_n$$

where $Q_{meas,n}$ again represents the level of the signal received on the beacon channel.

8.4.3 Intersystem Handover

In this section we describe the procedures used by the network and the mobile to handle mobility between GSM and UMTS systems. In the examples that follow, we will look at cases of mobility:

- when a packet-switched service is active;
- when a circuit-switched application is active.

Handover of PS Services

In GPRS networks, unlike GSM circuit-switched services, user mobility is the responsibility of the mobile, using a mechanism close to the cell reselection mechanism used in idle mode. When a decision to change cell is taken, the network does not strictly speaking carry out a handover; the mobile simply has to signal to the network the new cell to which it belongs.

The same principle has been adopted for the handover of PS services between GSM and UMTS.

Mobility from GPRS to UMTS

When a mobile in GPRS mode decides to select a UTRAN cell, it must carry out a change of location procedure, even if the old cell and the new cell belong to the same routing area. The procedure is identical to that described in the section on management of mobility in idle mode.

When the old cell and the new cell are not dependent on the same SGSN, a signalling exchange takes place between the SGSNs, allowing the new SGSN to obtain the attributes of the PDP context or contexts that were activated (Figure 8.28).

Once the PDP contexts have been transferred to the UMTS SGSN, it asks the RNC to allocate the corresponding resources, just as if a call were being set up.

Mobility from UMTS to GPRS

The principle is the same as before. The mobile, having selected a cell from the GSM network, must carry out a routing area update procedure. Once the PDP contexts have been transferred from the old SGSN to the new one, the resources used in the UTRAN are released.

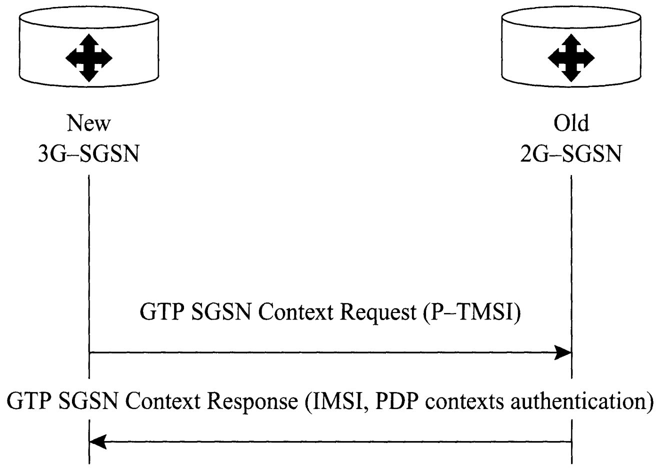


Figure 8.28 Transfer of PDP contexts from GSM to UMTS.

Handover of CS services

Handover from GSM to UMTS

In the example in Figure 8.29, a call has been set up on the GSM network. The GSM MSC is therefore the anchor MSC for the call, even after the mobile has moved over into the UTRAN access technology.

The handover from GSM to UMTS takes place in three phases:

1. *Preparation*. When the conditions for a handover to UMTS are satisfied, the BSC sends a *handover required* message, containing the identity of the target RNC. The request is transmitted to the target RNC through the UTRAN MSC, over the MAP-E and Iu interfaces. In the process, the parameters of the channel used by the mobile in the current GSM cell are translated into attributes of the RAB in the *relocation message*.

   The resource allocation phase, not shown in detail in Figure 8.29, is identical to that carried out during a hard handover. The GSM MSC is informed of the success of this phase by the MAP message *prepare handover ack*. Since it is the anchor MSC for the call, a circuit is set up with the new MSC, by a signalling exchange on the ISUP layer (see Section 7.2.7), *initial address message* and *address complete message*.
2. *Execution*. The *handover command* message contains the *handover* message sent to the mobile, which includes the description of the target resources in the UTRAN cell. This message will have been constructed by the target RNC and transferred to the BSC via the Iu, E and A interfaces. The message *relocation complete* marks the end of the changeover procedure from the point of view of the new RNC.
3. *Resource de-allocation*. The end of the handover is signalled to the GSM MSC, which must then release the resources used in the old BSS.

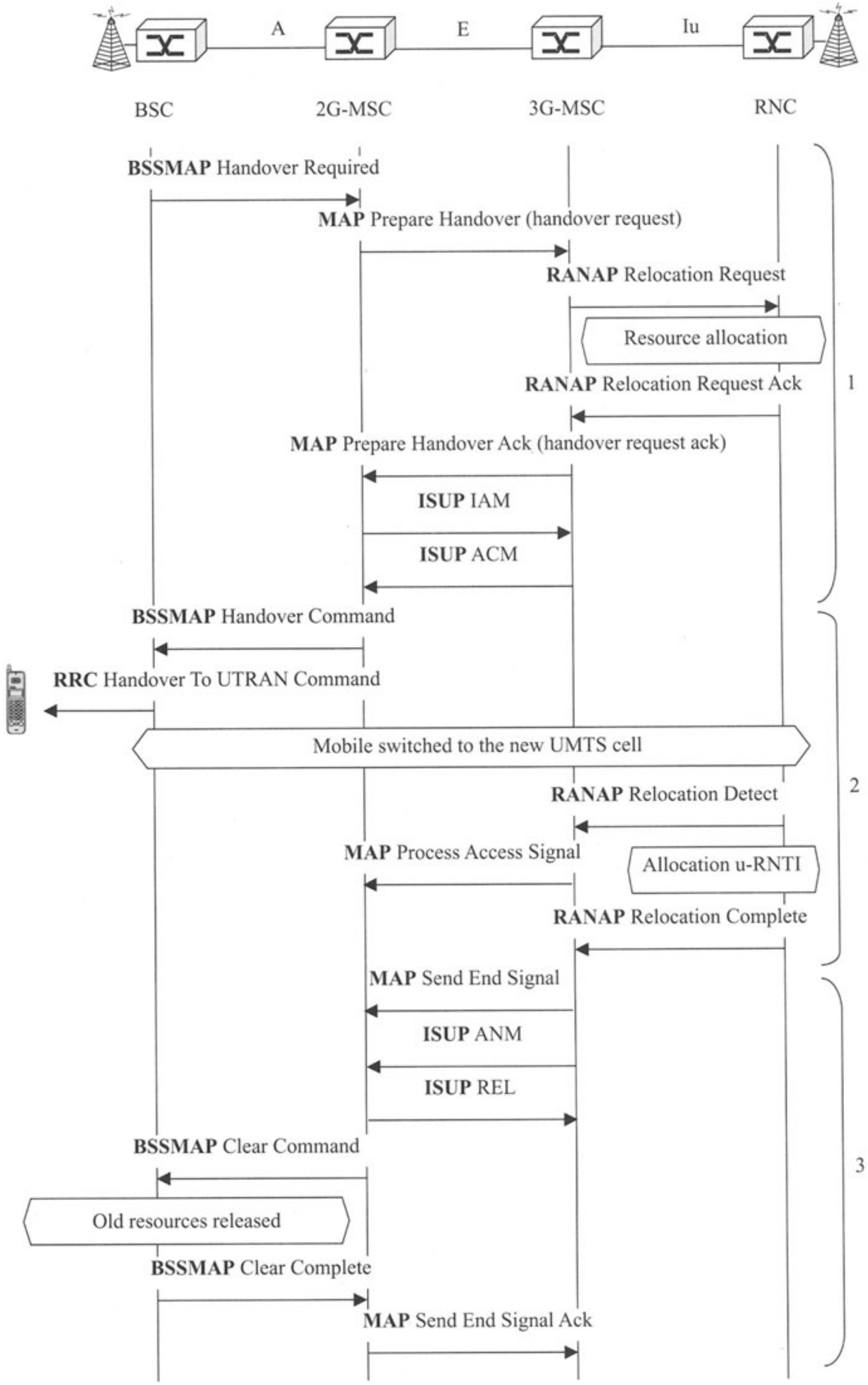


Figure 8.29 Handover from GSM to UMTS.

Handover from UMTS to GSM

The process of handing over from UMTS to GSM is quite similar to that described above. The same three phases are found as in handover from GSM to UMTS (Figure 8.30).

8.4.4 Security and Intersystem Handover

In Chapter 7, we saw the procedures provided in the UMTS network to authenticate users connected to the network and to protect the confidentiality of the data transmitted. GSM networks also include such procedures, following very similar ideas but using different algorithms and parameters (for example, encryption keys).

In order to avoid gaps in the security system (which would significantly reduce the effectiveness of the protection provided), it is necessary to guarantee continuity of encryption and integrity checking both to the user and to the network.

The two examples that follow show the mechanisms specified in the standard for maintaining the confidentiality of data exchanged between the user and the network in the case of intersystem handover for a circuit-switched service, be it voice or data.

Handover from GSM to UMTS

The procedure is shown in Figure 8.31.

When the mobile is active in a GMS cell, encryption is implemented using a key K_c and an algorithm (A5.1 or A5.2) specific to the GSM standard.

During the handover from a GSM cell, the UTRAN MSC uses the key K_c and conversion algorithms C4 and C5 to derive the encryption and integrity keys, CK and IK, used in the UTRAN cells. Keys CK and IK are then transmitted to the target RNS, which will proceed to implement integrity checking and encrypted transmission when the mobile starts to use the resources of the new UMTS cell, by means of the algorithms specific to UMTS.

For its part, the mobile must also convert the keys using algorithms C4 and C5.

Handover from UMTS to GSM

The procedure is shown in Figure 8.32.

The procedure used for handover from UMTS to GSM follows an analogous principle. The encryption key for the GSM network, K_c, is obtained by using algorithm C3 on the keys IK and CK. As in the previous case, the change is carried out simultaneously in the mobile and in the UMTS MSC.

8.5 Radio Measurements

The management of radio measurements is an inescapable feature of cellular communication networks. We have seen in the preceding sections that all the mobility procedures, whether in idle mode (i.e. cell reselection) or in connected mode, rely on physical measurements related to the radio interface.

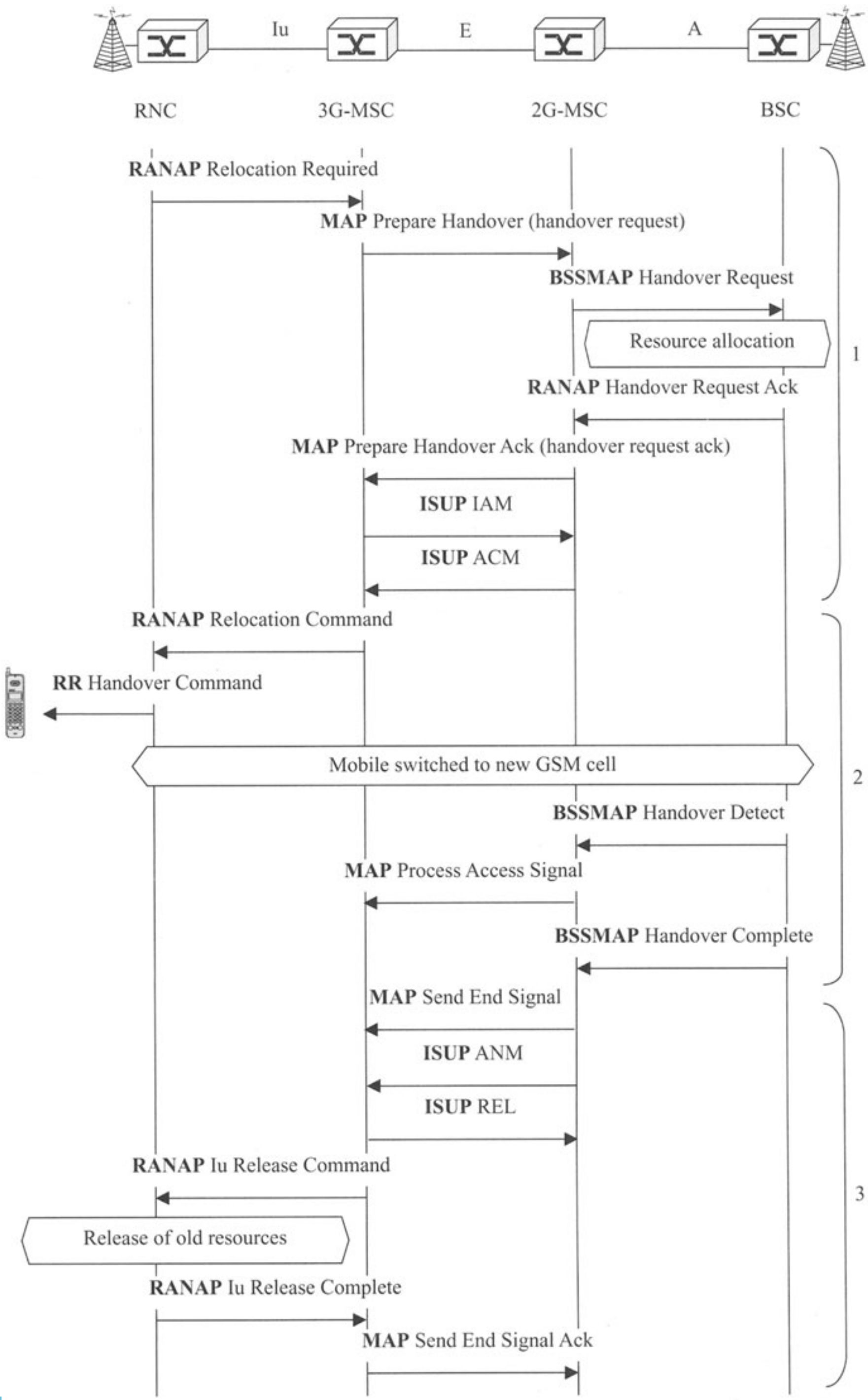


Figure 8.30 Handover from UMTS to GSM.

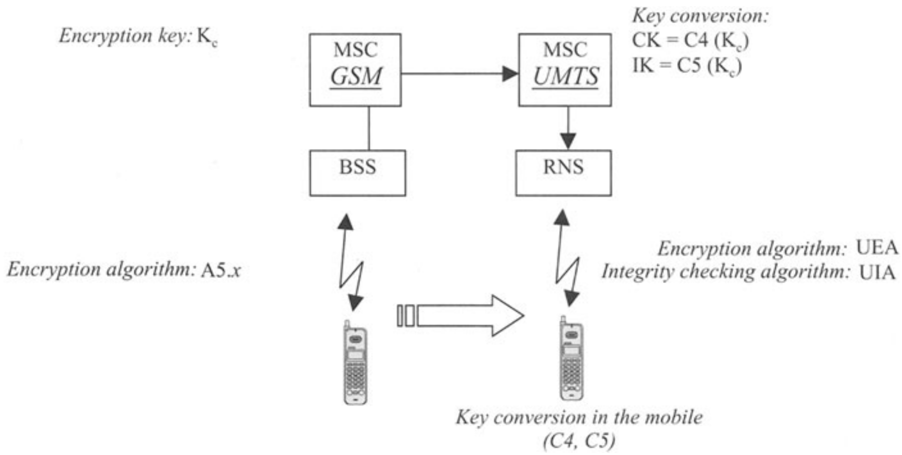


Figure 8.31 Handover from GSM to UMTS

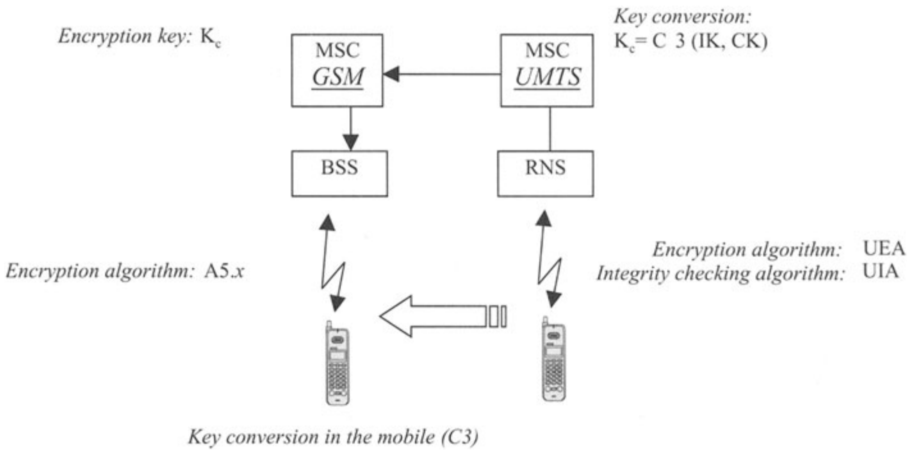


Figure 8.32 Handover from UMTS to GSM.

In this area, the UMTS standard defines a mechanism that is very rich and flexible in comparison with what was available in GSM. The following sections give an overview of what the standard provides and describe the procedure for compressed mode transmission, used to allow mobiles to carry out simultaneous measurements on different frequencies or different radio technologies.

8.5.1 Measurement Configuration

So far as measurement configuration is concerned, the line taken by UMTS is one of complete flexibility. In contrast to GSM, the standard makes no a priori requirements. It is up to the network, and hence to the implementer or the operator, to define the following items:

- the physical quantities that the mobile is required to measure;
- the pre-processing that the mobile must carry out on the raw data;
- the transmission mode used by the mobile to transmit the measurements back to the network.

Physical Quantities to be Measured

The UMTS standard defines a wide variety of quantities that the mobile might measure. Among the more useful, we might mention the following:

- CPICH RSCP (*Received Signal Code Power*), which is the power received and measured by the mobile on the common pilot channel (CPICH) in a UTRAN/FDD cell;
- CPICH E_c / N_o , the energy received by the mobile per chip on the CPICH of a UTRAN/FDD cell divided by the power density in the frequency band;
- GSM carrier RSSI, the signal level received by the mobile on the beacon channel of a neighbouring GSM cell;
- UE Tx, the total power transmitted by the mobile on a carrier frequency.

During the measurement configuration phase, the network can ask the mobile to carry out one or more of these measurements. In addition, the network must give the mobile a list of the objects on which the measurements are to be made, i.e. a list of neighbouring cells, whether using UTRAN or some other technology such as GSM.

Pre-processing the Raw Data

In order to reduce the computation time taken up in processing the measurements, the network may ask the mobile to pre-process the raw data. This pre-processing is also known as layer 3 filtering.

The filtering algorithm applied by the module is:

$$F_n = (1 - \alpha) \cdot F_{n-1} + \alpha \cdot M_n$$

where

- F_n is the filtered value at time n,
- M_n is the raw measurement at time n, and
- α is the filtering coefficient.

An example of filtering carried out with different values of α is shown in Figure 8.33.

This form of filtering is equivalent to a low pass filter. (The formula given above is, in fact, the digital form of an analogue filter of RC type.) The purpose of this type of filter is to eliminate short-term variations in the signal while preserving the main trends. In practical terms, layer 3 filtering can prevent the network from making unjustified handover decisions.

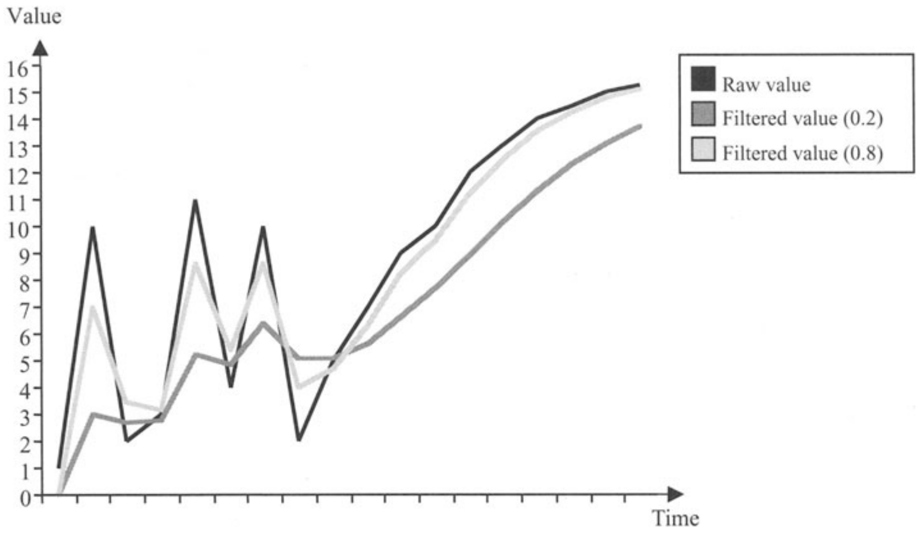


Figure 8.33 Example of level 3 filtering.

The norm specifies two ways in which the measurements can be transmitted back to the network: 'periodic' mode and 'triggered' mode.

In the periodic case, the mobile regularly transmits a measurement report to the network by means of the *RRC Measurement Report* message. The period, between 250 ms and 64 s, is set by the network. The network can also ask the mobile for a fixed number (from 1 to 64) of reports, or for an indefinite number.

In the triggered case, the report is only sent when the measurements satisfy certain conditions. The idea behind this mode is to reduce the number of signalling exchanges over the radio interface and the calculation time for the network nodes without in any way reducing the responsiveness of the network for mobility management.

The UMTS standard describes a large number of events that can trigger the sending of a report. Their precise formulation depends on the physical quantity being measured. Figure 8.34 shows an example of a triggered report based on the measurement of CPICH E_c/N_o .

In this example, cells 1, 2 and 3 form part of the active set (that is the set of cells that could participate in a soft handover) and cell 4 does not. The CPICH E_c/N_o measurement for cell 4 will only be reported by the mobile if it exceeds that of the worst cell in the active set (cell 3). Once the measurement has been received, the network can decide whether or not to add cell 4 to the active set, or even replace cell 3 in the active set by cell 4.

Measurements Performed by the Node B

In a way analogous to the mobile, the Node B can also perform a certain number of measurements at the request of the RNC that controls it.

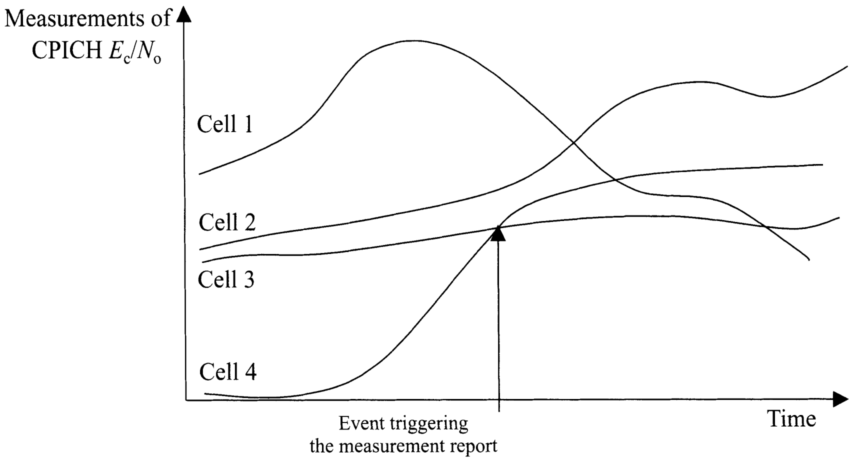


Figure 8.34 Example of measurement report triggered by an event.

Among the possible measurements envisaged by the UTRAN standard, there are, for example:

- measurement of the average bit error rate per transmission time interval, carried out by the Node B on one or more uplink transport channels;
- measurement of the number of PRACH headers acknowledged per 20 ms access slot (see Section 7.2.4).

For the same reasons (economising on computation time and on signalling traffic on the Iub interface), the Node B handles the measurements in much the same way as the mobile. At the request of the RNC, the Node B can filter the measurements using the same algorithm as that defined for the mobile. Once the measurement has been filtered, the Node B transmits the report either in periodic mode or in triggered mode.

General View

Figure 8.35 shows a general view of the stages involved in an example of configuring and managing measurements in the access network.

Following call set-up, the RNC configures the measurements that are to be carried out by the mobile and the Node B, using the RRC and NBAP protocols.

After layer 3 filtering has been applied, the measurements are transmitted to the RNC by the mobile, in triggered mode, and by the Node B, in periodic mode.

8.5.2 Compressed Mode

In the chapter on the UTRAN radio interface, we saw that transmission in FDD mode on the UTRAN radio interface was continuous. The transmission uses a sequence of 10 ms frames in both uplink and downlink directions.

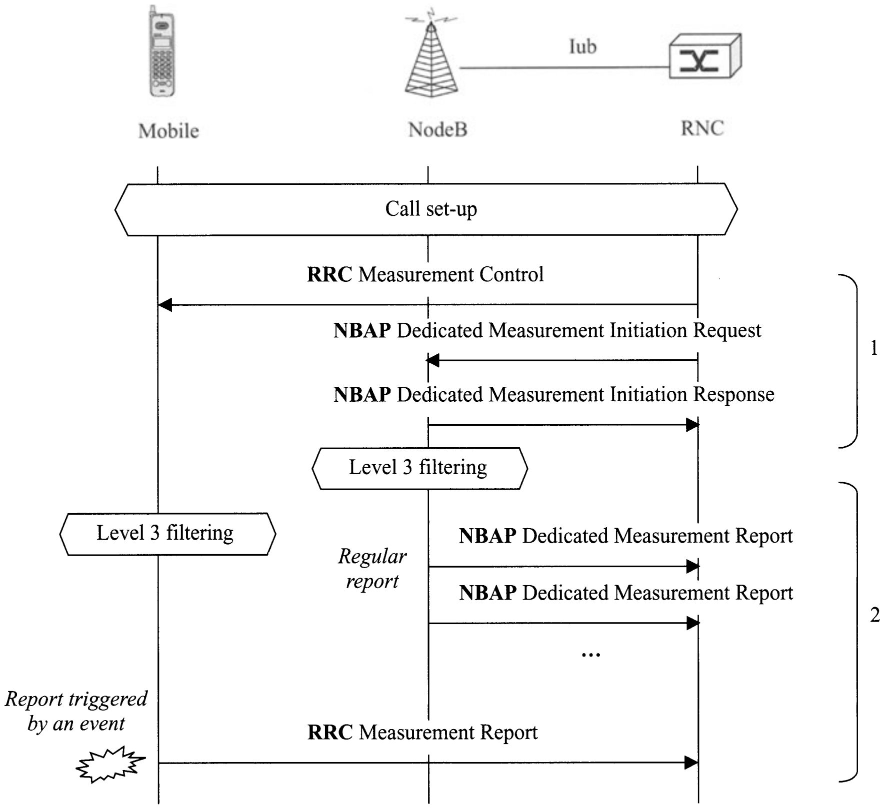


Figure 8.35 Example of measurement configuration.

In a single-frequency CDMA network (i.e. a network using only one carrier frequency), the mobile is simultaneously capable of receiving information sent by the network and carrying out measurements on the beacon channels of neighbouring CDMA cells, without any special constraint on the physical architecture.

The situation is not at all the same in a mixed network made up of cells using different access technologies (for example, UTRAN/FDD and GSM) or different CDMA carrier frequencies. In this case there are two solutions:

- The mobile is equipped with a dual receiver; it is thus able, for example, to carry out measurements on the neighbouring GSM cells using one channel, while receiving data on a UTRAN/FDD channel. This solution has the disadvantage of incurring additional cost at the terminal level, since a certain number of components have to be duplicated.
- The network provides gaps in the transmission of data. A mobile engaged in a call in a UTRAN/FDD cell can use these gaps to carry out measurements on the neighbouring GSM cells. This special transmission mode is known as compressed mode.

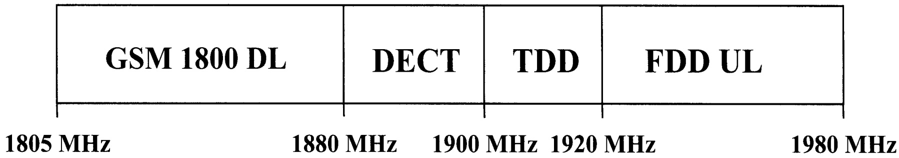


Figure 8.36 GSM 1800 and IMT-2000 frequency bands.

Independently of the problems of reception by the mobile, it may be necessary to resort to compressed mode in the uplink direction when the frequencies measured by the mobile on the downlink are close to the frequencies used on the uplink. This scenario can occur when a mobile is handling a call on an FDD frequency and is required to measure the signal level received on a GSM beacon channel in the 1800 MHz range (Figure 8.36).

The principle of compressed mode transmission is to create free spaces in the transmission, of a size equal to one or more time slots per frame. The length of a space can be up to seven time slots per frame, a frame consisting of 15 slots (example (a) in Figure 8.37). By combining two such spaces, one situated at the end of one frame and the other at the start of the next, it is possible to create spaces of a maximum of 14 time slots (example (b) in Figure 8.37), which is almost equal to the length of an FDD frame ($14 \times 10/15 = 9.33$ ms).

For a mobile in UMTS mode, carrying out measurements on a GSM cell implies decoding two sub-channels broadcast on the GSM cell's beacon channel:

- the frequency correction channel (FCCH);
- the synchronisation channel (SCH), which broadcasts the BSIC associated with the cell.

These channels are broadcast at time intervals fixed by the GSM standard, based on a multi-frame of length 51, representing:

$$51 \times 8 \times 15/26 = 235.4 \text{ ms, approximately.}$$

Since the broadcasting of UTRAN/FDD frames and GSM frames is not synchronised, it is impossible to predict the precise time at which the UTRAN will send a compressed frame that would correspond to the time at which the FCCH and SCH will be broadcast in the GSM cells.

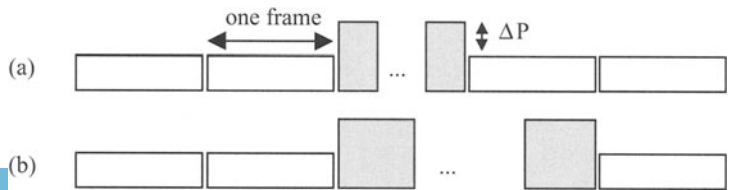


Figure 8.37 Two examples of compressed frames.

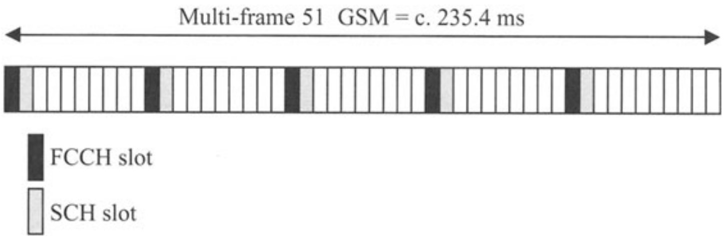


Figure 8.38 Complex mode and the compressed frame.

Because of the length of the time slot in the GSM frame (15/26 = 0.577 ms), the length of the multi-frame (51 time slots) and of the FDD frame (10 ms) are not multiples of each other. The transmission gaps in the FDD frame will slide relative to the structure of the GSM beacon channel multi-frame and in places coincide with the time at which the FCCH and SCH are broadcast, thus allowing the mobile to read those channels. (See Figure 8.38.)

To reinforce this sliding mechanism and reduce the time taken to acquire the measurements, compressed mode transmission is composed of an alternating sequence of two patterns of different lengths, TGPL1 and TGPL2 (TGPL: *Transmission Gap Pattern Length)*, expressed in units of FDD frames. This is shown in Figure 8.39.

Each pattern contains two transmission gaps, of length TGL1 and TGL2 slots (TGL: *Transmission Gap Length*), with a distance of TGD slots (TGD: *Transmission Gap Distance*) between them.

TGSN (*Transmission Gap Starting Slot Number*) is the distance in slots between the start of the first frame of the pattern and the start of the first transmission gap. TGPRC (*Transmission Gap Pattern Repetition Count*) is the number of patterns used in the compressed mode transmission sequence.

The UTRAN standard offers several different methods for compressing the data sent in a frame:

- 'Puncturing': after channel coding has been applied, a certain number of bits in the frame are removed. This method reduces the redundancy resulting from the coding of the data. The compressed frame is thus more sensitive to transmission errors.
- SF (spreading factor) reduction: the compressed frame is transmitted with an OVSF code whose spreading factor is half that of a normal frame.
- Higher layer scheduling: consists in using a specific frame format in order to decrease the number of useful bits transmitted.

In the first two cases, data compression is obtained at the expense of the protection offered to user data. In order not to impair the communication quality, it is

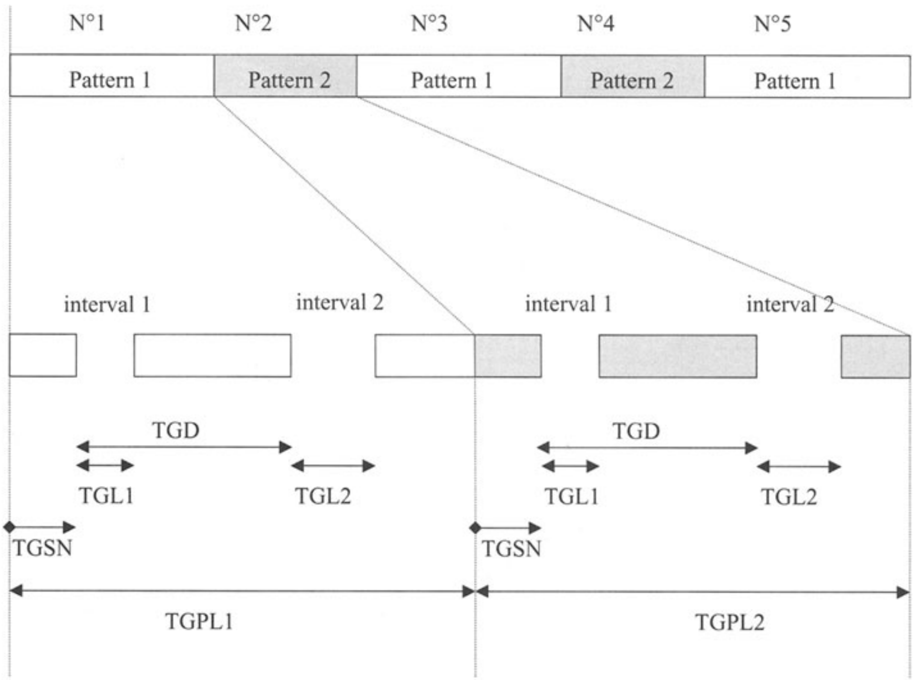


Figure 8.39 Compressed mode parameters.

necessary to increase the transmission power for the duration of the compressed frame.

Figure 8.40 shows the procedures employed in compressed mode transmission and the equipment involved.

1 Before compressed mode transmission can take place, there must first be a configuration phase in which the network provides the compressed mode transmission parameters to the mobile and the Node B. In the example in Figure 8.40, this phase takes place during the resource allocation phase of the call (on receipt of the RAB assignment request). Configuration includes setting all the parameters shown in Figure 8.39.
2 The configuration phase is followed by the activation of compressed mode transmission. The decision to enter compressed mode is taken by the SRNC. Compressed mode transmission is then activated at the level of the mobile (by means of a *Measurement Control* message in the RRC protocol) and at the level of the Node B (by means of the *Compressed Mode Command* in the NBAP protocol).

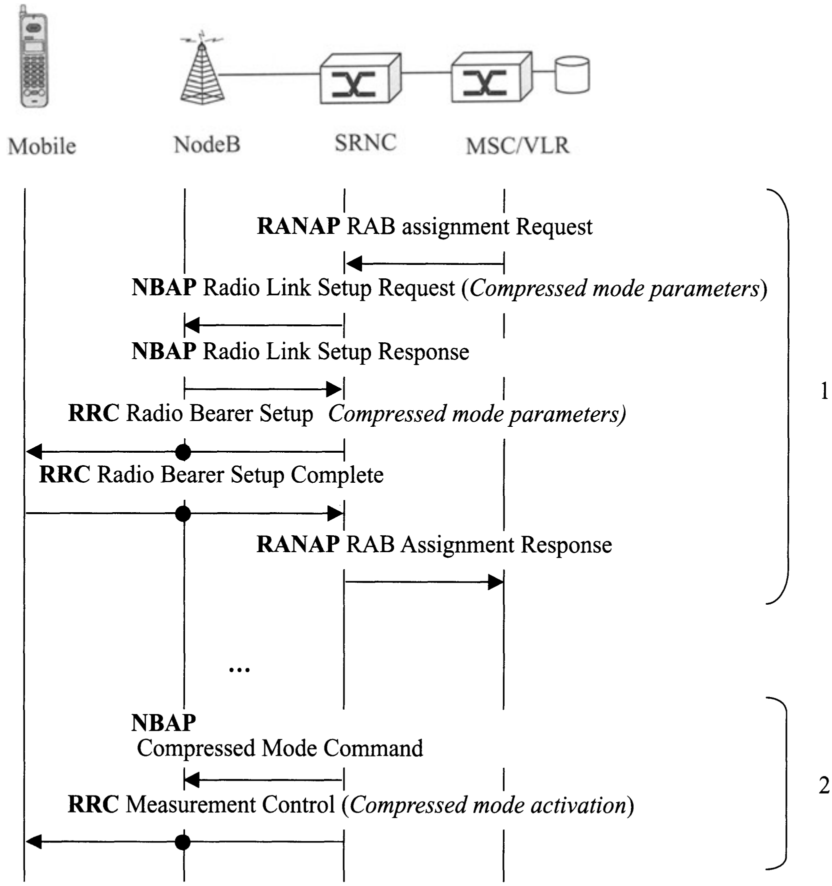


Figure 8.40 Configuration and activation of compressed mode.

Appendix A

UMTS Specifications

This appendix contains a list of the SGPP specifications for UMTS as well as describing the role of the various series. All these specifications are available on the 3GPP's public web site at http://www.3gpp.org.

A.1 Introduction

The UMTS standard includes a large number of documents (more than 300) that are continually evolving as new services or improvements are brought into the standard. In order to allow the industry to develop the technology on a base of stable specifications, the UMTS standard is made up of different versions.

Figure A.1 shows the principal versions of both the GSM and the UMTS standards. Phase 1 of the GSM standard was the first version available. Phase 2+ includes most of the important developments in the standard, such as multi-band networks (900 MHz/1800 MHz), GPRS and HSCSD.

Since the last version, which appeared in 1999, no important features have been added to the GSM standard. Given the emergence of UMTS networks, it is unlikely that the GSM standard will develop after 2001. In contrast, non-access stratum specifications, common to second and third generation networks will continue to evolve within the UMTS framework.

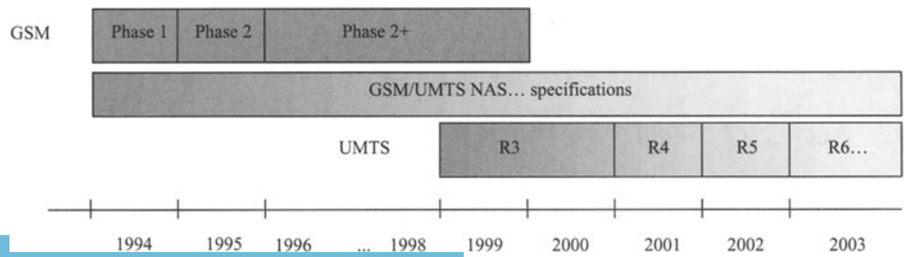


Figure A.1 Versions of the standard.

Table A.1 Principal changes in version 4.

	Description
Multi-media messaging service	MMS is an extension of the short message service, offering the user the possibility of transmitting audio messages (voice, music), images, video, etc.
Tandem-free operation Transcoder-free operation	Improvements aimed at reducing the use of speech transcoding for mobile to mobile calls.
UTRAN transport evolution	A set of functions aimed at optimising the use of ATM/AAL2 in the UTRAN transport network.

Table A.2 Principal changes in version 5.

	Description
High speed downlink packet access (HSDPA)	Improves the data rate and transmission delay for data transport on a downlink shared channel.
UTRAN transport evolution	Introduction of IP into the UTRAN transport network, as an alternative to the ATM/AAL2 layers.
IP-based multimedia services	Migration of the core network to an all-IP architecture for the support of user applications (voice, data transmission, etc.).
Wideband AMR	Improvement in the quality of speech, thanks to the use of coder/decoders allowing a larger bandwidth (7 kHz) to be transmitted.

Version 3 (release 99) constitutes the first operational version of the UMTS standard. It includes all the basic functions that are needed to develop and implement a UMTS network. Versions 4 and 5, released respectively in mid-2001 and mid-2002, are based on version 3. They include a number of improvements and technological developments whose principles are summarised in Tables A.1 and A.2.

A.2 Series 21

This series contains general documents concerning, for example, methods of working in the committees and procedures for document management.

21.900 3GPP Working Methods
21.904 UE Capability Requirements (UCR)
21.905 3G Vocabulary
21.910 Multi-Mode UE issues
21.978 Feasibility Technical Report – CAMEL Control of VoIP Services

A.3 Series 22: Service Definition

This series contains documents describing and specifying all the UMTS services. One of the UMTS priorities is to achieve maximum compatibility, in terms of services, with second generation systems. It is in this series, therefore, that the majority of the documents inherited from GSM, describing services common to

UMTS and GSM, will be found, including, for example, call barring and call forwarding. Other services specific to UMTS, such as multi-call have been added.

22.001 Principles of Telecommunication Services Supported by a GSM Public Land Mobile Network (PLMN)
22.002 Bearer Services Supported by a GSM PLMN
22.003 Tele-services Supported by a GSM Public Land Mobile Network (PLMN)
22.004 General on Supplementary Services
22.011 Service Accessibility
22.016 International Mobile Equipment Identities (IMEI)
22.022 Personalisation of GSM ME Mobile Functionality Specification – Stage 1
22.024 Description of Charge Advice Information (CAI)
22.030 Man–Machine Interface (MMI) of the Mobile Station (MS)
22.034 High Speed Circuit Switched Data (HSCSD) – Stage 1
22.038 SIM Application Toolkit (SAT) – Stage 1
22.041 Operator Determined Call Barring
22.042 Network Identity and Time Zone (NITZ) – Stage 1
22.043 Support of Localised Service Area (SoLSA) – Stage 1
22.053 Tandem Free Operation of Speech Codecs – Stage 1 Service Description
22.057 Mobile Station Application Execution Environment (MExE) – Stage 1
22.060 General Packet Radio Service (GPRS) – Stage 1
22.066 Support of Mobile Number Portability (MNP) – Stage 1
22.067 Enhanced Multi-level Precedence and Pre-emption service (eMLPP) – Stage 1
22.071 Location Services (LCS) – Stage 1 (T1P1)
22.072 Call Deflection (CD) – Stage 1
22.078 CAMEL Phase 3 – Stage 1
22.079 Support of Optimal Routing – Stage 1
22.081 Line Identification Supplementary Services – Stage 1
22.082 Call Forwarding (CF) Supplementary Services – Stage 1
22.083 Call Waiting (CW) and Call Hold (HOLD) Supplementary Services – Stage 1
22.084 Multi-Party (MPTY) Supplementary Service – Stage 1
22.085 Closed User Group (CUG) Supplementary Services – Stage 1
22.086 Advice of Charge (AoC) Supplementary Services – Stage 1
22.087 User-to-User Signalling (UUS) – Stage 1
22.088 Call Barring (CB) Supplementary Services – Stage 1
22.090 Unstructured Supplementary Service Data (USSD) – Stage 1
22.091 Explicit Call Transfer (ECT) Supplementary Service – Stage 1
22.093 Call Completion to Busy Subscriber (CCBS) – Stage 1
22.094 Follow Me Stage 1
22.096 Calling Name Presentation (CNAP) – Stage 1 (T1P1)
22.097 Multiple Subscriber Profile (MSP) – Stage 1
22.100 UMTS Phase 1
22.101 UMTS Service Principles
22.105 Services & Service Capabilities
22.115 Service Aspects Charging and Billing
22.121 Provision of Services in UMTS – The Virtual Home Environment
22.129 Handover Requirements between UMTS and GSM or Other Radio Systems
22.135 Multicall Stage1
22.140 Multimedia Messaging Service Stage 1
22.945 Study of Provision of Fax Service in GSM and UMTS

22.960 Mobile Multimedia Services
22.970 Virtual Home Environment Report
22.971 Automatic Establishment of Roaming Relationships
22.975 Advanced Addressing

A.4 Series 23: Service Implementation

This series contains the documents that specify how the services specified in Series 22 are implemented. Complementing the purely functional view taken in Series 22, these documents specify the functions of the nodes and other UMTS network equipment.

23.002 Network Architecture
23.003 Numbering, Addressing and Identification
23.007 Restoration Procedures
23.008 Organisation of Subscriber Data
23.009 Handover Procedures
23.011 Technical Realization of Supplementary Services – General Aspects
23.012 Location Registration Procedures
23.014 Support of Dual Tone Multi-Frequency (DTMF) Signalling
23.015 Technical Realization of Operator Determined Barring (ODB)
23.016 Subscriber Data Management – Stage 2
23.018 Basic Call Handling – Technical Realisation
23.032 Universal Geographical Area Description (GAD)
23.034 High Speed Circuit Switched Data (HSCSD) – Stage 2
23.038 Alphabets & Language
23.039 Interface Protocols for the Connection of Short Message Service Centers (SMSCs) to Short Message Entities (SMEs)
23.040 Technical Realization of SMS Point to Point
23.041 Technical Realization of Short Message Service Cell Broadcast (SMSCB)
23.042 Compression Algorithm for SMS
23.054 Shared Interworking Functions – Stage 2
23.057 Mobile Station Application Execution Environment (MExE)
23.060 General Packet Radio Service (GPRS) Service Description – Stage 2
23.066 Support of GSM Mobile Number Portability (MNP) – Stage 2
23.067 Enhanced Multi-level Precedence and Pre-emption Service (EMLPP) – Stage 2
23.072 Call Deflection Supplementary Service – Stage 2
23.073 Support of Localised Service Area (SoLSA) – Stage 2
23.078 CAMEL Stage 2
23.079 Support of Optical Routing, Phase 1 – Stage 2
23.081 Line Identification Supplementary Services – Stage 2
23.082 Call Forwarding (CF) Supplementary Services – Stage 2
23.083 Call Waiting (CW) and Call Hold (HOLD) Supplementary Service – Stage 2
23.084 Multi-Party (MPTY) Supplementary Service – Stage 2
23.085 Closed User Group (CUG) Supplementary Service – Stage 2
23.086 Advice of Charge (AoC) Supplementary Service – Stage 2
23.087 User-to-User Signalling (UUS) – Stage 2
23.088 Call Barring (CB) Supplementary Service – Stage 2

23.090 Unstructured Supplementary Service Data (USSD) – Stage 2
23.091 Explicit Call Transfer (ECT) Supplementary Service – Stage 2
23.093 Call Completion to Busy Subscriber (CCBS) – Stage 2
23.094 Follow Me Stage 2
23.096 Name Identification Supplementary Service – Stage 2
23.097 Multiple Subscriber Profile (MSP) – Stage 2
23.101 General UMTS Architecture
23.107 Quality of Service, Concept and Architecture
23.108 Mobile Radio Interface Layer 3 Specification Core Network Protocols – Stage 2 (Structured Procedures)
23.110 UMTS Access Stratum Services and Functions
23.116 Super Charger – Stage 2
23.119 Gateway Location Register (GLR) – Stage 2
23.121 Architecture Requirements for Release 99
23.122 Non-access Stratum Functions Related to Mobile Station (MS) in Idle Mode
23.127 Virtual Home Environment/Open Service Architecture
23.135 Multicall – Stage 2
23.140 Multimedia Messaging Service (MMS) Functional Description – Stage 2
23.146 Technical Realisation of Facsimile Group 3 Service – Non-Transparent
23.153 Out of Band Transcoder Control – Stage 2
23.171 Functional Stage 2 Description of Location Services in UMTS
23.814 Separating RR and MM Specific Parts of the MS Classmark
23.821 Architecture Requirements for Release 2000
23.908 Technical Report on Pre-paging
23.909 Technical Report on the Gateway Location Register
23.910 Circuit Switched Data Bearer Services
23.911 Technical Report on Out-of-band Transcoder Control
23.912 Technical Report on Super-charger
23.923 Combined GSM and Mobile IP Mobility Handling in UMTS IP CN
23.930 Iu Principles
23.972 Multimedia Telephony

A.5 Series 24: UE/CN Protocols

This series contains descriptions of all the protocols used between the UMTS mobile and the core network. Among others, the following protocols are covered:

- 24.008: the MM, CC, GMM and SM protocols;
- 24.011: the protocols for the short message service.

24.002 GSM-UMTS Public Land Mobile Network Access Reference Configuration
24.007 Mobile Radio Interface Signalling Layer 3 – General Aspects
24.008 Mobile Radio Interface Layer 3 Specification Core Network Protocols – Stage 3
24.010 Mobile Radio Interface Layer 3 – Supplementary Services Specification – General Aspects
24.011 Point-to-Point (PP) Short Message Service (SMS) Support on Mobile Radio Interface
24.012 Short Message Service Cell Broadcast (SMSCB) Support on the Mobile Radio Interface
24.022 Radio Link Protocol (RLP) for Data and Telematic Services on the (MS-BSS) Interface and the Base Station System – Mobile-Services Switching Centre (BSS-MSC) Interface

24.030 Location Services LCS – Stage 3
24.067 Enhanced Multi-Level Precedence and Pre-emption Service (eMLPP) – Stage 3
24.072 Call Deflection Supplementary Service – Stage 3
24.080 Mobile Radio Layer 3 Supplementary Service Specification – Formats and Coding
24.081 Line Identification Supplementary Service – Stage 3
24.082 Call Forwarding Supplementary Service – Stage 3
24.083 Call Waiting (CW) and Call Hold (HOLD) Supplementary Service – Stage 3
24.084 Multi-Party (MPTY) Supplementary Service – Stage 3
24.085 Closed User Group (CUG) Supplementary Service – Stage 3
24.086 Advice of Charge (AoC) Supplementary Service – Stage 3
24.087 User-to-User Signalling (UUS) – Stage 3
24.088 Call Barring (CB) Supplementary Service – Stage 3
24.090 Unstructured Supplementary Service Data (USSD) – Stage 3
24.091 Explicit Call Transfer (ECT) Supplementary Service – Stage 3
24.093 Call Completion to Busy Subscriber (CCBS) – Stage 3
24.096 Name Identification Supplementary Service – Stage 3
24.135 Multicall – Stage 3

A.6 Series 25: Access Network

All the specifications relating to the UTRAN will be found in this series. Because of the great variety of documents contained in this series, they have been split into several sub-series, each treating some particular aspect of the UTRAN, as follows:

- 25.1*xx*: radio performance aspects;
- 25.2*xx*: level 1 of the radio interface, split as follows:
  - 25.21*x*: FDD mode specifications;
  - 25.22*x*: TDD mode specifications;
- 25.3*xx*: the radio interface protocols, split as follows:
  - 5.32*x*: level 2 protocols;
  - 25.33*x*: UTRAN level 3 protocols;
- 25.4*xx*: UTRAN protocols and network interfaces, split into three sub-series:
  - 25.41*x*: specifications of the Iu interface;
  - 25.42*x*: specifications of the Iur interface;
  - 25.43*x*: specifications of the Iub interface.

25.101 UE Radio Transmission and Reception (FDD)
25.102 UE Radio Transmission and Reception (TDD)
25.103 RF Parameters in Support of RRM
25.104 BTS Radio Transmission and Reception (FDD)
25.105 BTS Radio Transmission and Reception (TDD)
25.113 BTS EMC
25.123 RF Parameters in Support of RRM (TDD)
25.133 RF Parameters in Support of RRM (FDD)
25.141 Base Station Conformance Testing (FDD)
25.142 Base Station Conformance Testing (TDD)
25.201 Physical Layer – General Description
25.211 Physical Channels and Mapping of Transport Channels onto Physical Channels (FDD)

25.212 Multiplexing and Channel Coding (FDD)
25.213 Spreading and Modulation (FDD)
25.214 FDD Physical Layer Procedures
25.215 Physical Layer Measurements (FDD)
25.221 Physical Channels and Mapping of Transport Channels onto Physical Channels (TDD)
25.222 Multiplexing and Channel Coding (TDD)
25.223 Spreading and Modulation (TDD)
25.224 TDD Physical Layer Procedures
25.225 Physical Layer Measurements (TDD)
25.301 Radio Interface Protocol Architecture
25.302 Services Provided by the Physical Layer
25.303 Interlayer Procedures in Connected Mode
25.304 UE Procedures in Idle Mode and Procedures for Cell Reselection in Connected Mode
25.305 Stage 2 Functional Specification of Location Services in UTRAN (LCS)
25.306 UE Radio Access capabilities definition
25.321 MAC Protocol Specification
25.322 RLC Protocol Specification
25.323 Packet Data Convergence Protocol (PDCP)
25.324 Radio Interface for Broadcast/Multicast Services
25.331 RRC Protocol Specification
25.401 UTRAN Overall Description
25.402 Synchronisation in UTRAN Stage 2
25.410 UTRAN Iu Interface – General Aspects and Principles
25.411 UTRAN Iu Interface Layer 1
25.412 UTRAN Iu Interface Signalling Transport
25.413 UTRAN Iu Interface RANAP Signalling
25.414 UTRAN Iu Interface Data Transport & Transport Signalling
25.415 UTRAN Iu Interface User Plane Protocols
25.419 UTRAN Iu Interface – Cell Broadcast Protocols between SMS-CBC and RNC
25.420 UTRAN Iur Interface – General Aspects and Principles
25.421 UTRAN Iur Interface Layer 1
25.422 UTRAN Iur Interface Signalling Transport
25.423 UTRAN Iur Interface RNSAP Signalling
25.424 UTRAN Iur Interface Data Transport & Transport Signalling for CCH Data Streams
25.425 UTRAN Iur Interface User Plane Protocols for CCH Data Streams
25.426 UTRAN Iur and Iub Interface Data Transport & Transport Signalling for DCH Data Streams
25.427 UTRAN Iur and Iub Interface user Plane Protocols for DCH Data Streams
25.430 UTRAN Iub Interface – General Aspects and Principles
25.431 UTRAN Iub Interface Layer 1
25.432 UTRAN Iub Interface Signalling Transport
25.433 UTRAN Iub Interface NBAP Signalling
25.434 UTRAN Iub Interface Data Transport & Transport Signalling for CCH Data Streams
25.435 UTRAN Iub Interface user Plane Protocols for CCH Data Streams
25.442 UTRAN Implementation Specific O & M Transport
25.832 Manifestations of Handover and SRNS Relocation
25.921 Guidelines and Principles for Protocol Description and Error Handling
25.922 Radio Resource Management Strategies
25.925 Radio Interface for Broadcast/Multicast Services
25.931 UTRAN Functions, Examples of Signalling Procedures

25.941 RAN WG4 Document Structure
25.944 Channel Coding and Multiplexing Examples
25.990 Vocabulary for UTRAN

A.7 Series 26: Coders and Decoders

This series contains the specifications of the coders and decoders used in UMTS networks, in particular for telephonic applications.

26.071 AMR Speech Codec – General Description
26.073 AMR Speech Codec – C-Source Code
26.074 AMR Speech Codec – Test Sequences
26.090 AMR Speech Codec – Transcoding Functions
26.091 AMR Speech Codec – Error Concealment of Lost Frames
26.092 AMR Speech Codec – Comfort Noise for AMR Speech Traffic Channels
26.093 AMR Speech Codec – Source Controlled Rate Operation
26.094 AMR Speech Codec – Voice Activity Detector for AMR Speech Traffic Channels
26.101 AMR Speech Codec – Frame Structure
26.102 AMR Speech Codec – Interface to Iu and Uu
26.103 Codec lists
26.104 AMR Speech Codec – Floating point C-Code
26.110 Codec for Circuit Switched Multimedia Telephony Service – General Description
26.111 Codec for Circuit Switched Multimedia Telephony Service – Modifications to H.324
26.131 Narrow Band (3.1 kHz) Speech & Video Telephony Terminal Acoustic Characteristics
26.132 Narrow Band (3.1 kHz) Speech & Video Telephony Terminal Acoustic Test Specification
26.911 Codec for Circuit Switched Multimedia Telephony Service – Terminal Implementer's Guide
26.912 Codec for Circuit Switched Multimedia Telephony Service – Quantitative Performance Evaluation of H.324 Annex C over 3G
26.915 Transmission Planning Aspects of the Services in 3G PLMN System

A.8 Series 27: Data Applications

This series specifies the UMTS network functions intended specifically for the support of data applications.

27.001 General on Terminal Adaptation Functions (TAF) for Mobile Stations (MS)
27.002 Terminal Adaptation Functions (TAF) for Services using Asynchronous Bearer Capabilities
27.003 Terminal Adaptation Functions (TAF) for Services using Synchronous Bearer Capabilities
27.005 Use of Data Terminal Equipment – Data Circuit Terminating Equipment (DTE-DCE) Interface for Short Message Service (SMS) and Cell Broadcast Service (CBS)
27.007 AT Command Set for 3G User Equipment (UE)
27.010 Terminal Equipment to User Equipment (TE-UE) Multiplexer Protocol User Equipment (UE)

27.060 GPRS Mobile Stations Supporting GPRS
27.103 Wide Area Network Synchronisation
27.901 Report on Terminal Interfaces – An Overview
27.903 Discussion of Synchronisation Standards

A.9 Series 29: Core Network Protocols

This series contains the specifications of the protocols used in the UMTS core network and, in particular, the MAP and GTP protocols used for managing subscribers and circuit- and packet-switched calls.

29.002 Mobile Application Part (MAP)
29.007 General Requirements on Interworking between the PLMN and the ISDN or PSTN
29.010 Information Element Mapping between Mobile Station – Base Station System (MS-BSS) and Base Station System – Mobile-Services Switching Centre (BSS-MCS) Signalling Procedures and the Mobile Application Part (MAP)
29.011 Signalling Interworking for Supplementary Services
29.013 Signalling Interworking between ISDN Supplementary Services Application Service Element (ASE) and Mobile Application Part (MAP) Protocols
29.016 Serving GPRS Support Mode SGSN – Visitors Location Register (VLR) – Gs Interface Network Service Specification
29.018 Serving GPRS Support Mode SGSN – Visitors Location Register (VLR) – Gs Interface Layer 3 Specification
29.060 GPRS Tunnelling Protocol (GPT) across the Gn and Gp Interface
29.061 General Packet Radio Service (GPRS) – Interworking between the Public Land Mobile Network (PLMN) Supporting GPRS and Packet
29.078 CAMEL Phase 3 – Stage 3
29.108 Application of the Radio Access Network Application Part on the E-interface
29.119 GPRS Tunnelling Protocol (GTP) Specification for Gateway Location Register (GLR)
29.120 Mobile Application Part (MAP) Specification for Gateway Location Register (GLR) – Stage 3
29.198 Open Services Architecture API – Part 1
29.998 Open Services Architecture API – Part 2

A.10 Series 31: UIM Aspects

This series contains the specifications of the user identity module (UIM) and of the interfaces between the UIM and the external environment.

31.101 UICC/Terminal Interface – Physical and Logical Characteristics
31.102 Characteristics of the USIM Application
31.110 Numbering System for Telecommunication IC Card Applications
31.111 USIM Application Toolkit (USAT)
31.120 Terminal Tests for the UICC Interface – Part 1
31.121 Terminal Tests for the UICC Interface – Part 2
31.122 UICC Test Specification

A.11 Series 32: Operations and Maintenance

This series contains specifications of the interfaces and procedures for maintaining and administering UMTS networks.

32.005 GSM Charging CS Domain
32.015 GSM Charging PS Domain
32.101 3G Telecom Management Principles and High Level Requirements
32.102 3G Telecom Management Architecture
32.104 3G Performance Management
32.105 3G Charging Call Event Data
32.106 Configuration Management
32.111 Fault Management

A.12 Series 33: Security

This series contains the specifications of the security mechanisms in UMTS networks, including encryption and integrity checking.

33.102 Security Architecture
33.103 Security Integration Guidelines
33.105 Cryptographic Algorithm Requirements
33.106 Lawful Interception Requirements
33.107 Lawful Interception Architecture and Functions
33.120 Security Objectives and Principles
33.901 Criteria for Cryptographic Algorithm Design Process
33.902 Formal Analysis of the 3G Authentication Protocol with Modified Sequence Number Management
33.908 General Report on the Design, Specification and Evaluation of 3GPP Standard Confidentiality and Integrity Algorithms

A.13 Series 34: Testing

This series brings together all the test specifications for UMTS equipment necessary for conformance testing.

34.108 Common Test Environments for User Equipment (UE) Conformance Testing
34.109 Logical Test Interface (TDD and FDD)
34.121 Terminal Conformance Specification – Radio Transmission and Reception (FDD)
34.122 Terminal Conformance Specification – Radio Transmission and Reception (TDD)
34.123-1 UE Conformance Specification, Part 1 – Conformance Specification
34.123-2 UE Conformance Specification, Part 2 – ICS
34.124 Electro-magnetic Compatibility (EMC) for Terminal Equipment – Stage 1
34.907 Report on Electrical Safety Requirements and Regulations
34.925 Specific Absorption Rate (SAR) Requirements and Regulations in Different Regions

A.14 Series 35: Security Algorithms

This series brings together the specifications of all the security algorithms (integrity checking, encryption and authentication) employed in UMTS networks. These algorithms are confidential and are only communicated by the standards organisations under licence.

35.201 Confidentiality and Integrity Algorithms – f8 and f9 specifications
35.202 Confidentiality and Integrity Algorithms – Kasumi Algorithm Specification
35.203 Confidentiality and Integrity Algorithms –Implementers' Test Data
35.204 Confidentiality and Integrity Algorithms – Design Conformance Test Data
35.205 General Report on the Design, Specification and Evaluation of the MILENAGE Algorithm Set
35.206 Key Generation Functions f1, f1*, f2, f3, f4 and f5* – Algorithm Specifications
35.207 Key Generation Functions f1, f1*, f2, f3, f4 and f5* – Implementers' Test Data
35.208 Key Generation Functions f1, f1*, f2, f3, f4 and f5* – Design Conformance Test Data

Glossary

3GPP *3rd Generation Partnership Project.* Consortium of manufacturers and operators responsible for the definition of the UMTS standard.

3GPP2 *3rd Generation Partnership Project 2.* Consortium of manufacturers and operators responsible for the definition of the cdma2000 standard.

AC *Access Class.* The concept of access class is common to UMTS and GSM. It allows access to the network to be restricted, for example in case of overloading.

Active set Set of radio links between the mobile and the network used in soft handover.

AICH *Acquisition Indicator Channel.* A physical downlink channel used by the network to acknowledge the access headers sent by the mobile on the PRACH.

AMPS *Advanced Mobile Phone Service.* A second generation cellular radio-communication system used mainly in North America.

AN *Access Network.* Subset of the UMTS network consisting of the elements connected with radio access.

ANSI-41 North American equivalent of the MAP. A standard that describes the signalling and the services (call set-up, mobility management, subscriber management, etc.) supported by the core network.

APN *Access Point Name.* In packet-switched mode, identifies the data network (Internet port) to which the user desires access.

AS *Access Stratum.* UMTS network layer containing the functions connected with radio access.

ASC *Access Service Class.* The concept of service class is used to segment the access capacity of the PRACH depending on the services required by the users of the network.

ATM *Asynchronous Transfer Mode.*

AuC *Authentication Center.* Component of the core GSM or UMTS network that is able to authenticate users.

BC *Billing Center.* Responsible for collecting and storing call detail records.

BCCH *Broadcast Control Channel.* Logical control channel used for broadcasting system information.

BCH *Broadcast Channel.* Shared uni-directional transport channel.

BMC *Broadcast/Multicast Control.* The layer of the UTRAN radio interface responsible for broadcasting messages.

BSS *Base Station Subsystem*. The GSM access network.

CC *Call Control*. Protocol in GSM and UMTS networks responsible for handling calls in circuit-switched mode.

CCCH *Common Control Channel*. Logical control channel used for transmitting signalling information to users not connected to the network.

CCPCH *Common Control Physical Channel*. Contains two sub-channels, P-CCPCH (primary CCPCH) and S-CCPCH (secondary CCPCH).

CDMA *Code Division Multiple Access*.

CDR *Call Detail Record*. A docket (electronic) containing the information needed for billing the mobile user for a service he has used.

CGF *Charging Gateway Function*. Network element belonging to the PS domain of the core network. Used as an intermediate storage place for call detail records.

CK *Ciphering Key*. Key used by the UTRAN for encrypting user data.

CN *Core Network*. Subset of the UMTS network containing those elements independent of the radio access.

CRC *Cyclic Redundancy Check*. Method used for detecting transmission errors. It is used both on the fixed UMTS interfaces and on the UTRAN radio interface.

CS domain Subset of the UMTS core network responsible for applications in circuit-switched mode.

CTCH *Common Traffic Channel*. Logical traffic channel used for the transmission of data to a group of mobiles.

DCCH *Dedicated Control Channel*. Bi-directional logical control channel used for transmitting signalling messages to and from users connected to the network.

DCH *Dedicated Channel*. A transport channel dedicated to a single user.

DPDCH *Dedicated Physical Data Channel*.

DTCH *Dedicated Traffic Channel*. Logical traffic channel used for transmitting user data for users connected to the network.

DSCH *Downlink Shared Channel*.

DSSS *Direct Spread Sequence Spectrum*. A technique used on the UTRAN radio interface.

EDGE *Enhanced Data Rate for GSM Evolution*. A new modulation technique introduced in phase 2+ of the GSM standard.

EIR *Equipment Identity Register*. GSM and UMTS core network database containing a black list of mobile terminals not allowed access.

FACH *Forward Access Channel*. A shared downlink transport channel.

FDD *Frequency Division Duplex*. Separation of uplink and downlink paths by the use of different frequencies.

FDMA *Frequency Division Multiple Access*. Multiple access mechanism based on frequency division.

FP *Frame Protocol*. Protocol used for transmitting user data over the fixed UTRAN interfaces (Iu, Iub and Iur).

GGSN Gateway GPRS Support Node. Component of GSM/GPRS and UMTS core networks used as a gateway between packet-switched networks (Internet) and the wireless network.

GMM *GPRS Mobility Management*. GSM/GPRS and UMTS protocol responsible for security and mobility (in the sense of the core network) aspects of packet-switched calls.

GMSC *Gateway MSC*. MSC used as a gateway between the RTCP and GSM or UMTS networks.

GMSK *Gaussian Minimum Shift Keying*. Modulation technique used on the radio interface of GSM networks.

GPRS *General Packet Radio Service*. GSM packet-switched data transmission service

GSM *Global System for Mobile Communications*.

GTP *GPRS Tunnelling Protocol*. Data encapsulation protocol for packet-switched services. Used in the GSM/GPRS and UMTS core networks.

Handover Mechanism for transferring a mobile/network connection from one radio resource (cell) to another. To techniques are used: *hard handover*, in which the old resource is abandoned before passing to the new one, and *soft handover*, in which the use of the old and new resources overlaps.

HLR *Home Location Register*. Core network database (both GSM and UMTS), containing account information relating to the network subscribers.

HPLMN *Home PLMN*. The PLMN of the operator with whom the subscriber has an account.

HSCSD *High Speed Circuit Switched Data*. High speed circuit-switched data transmission service in the GSM standard.

IK *Integrity Key*. Key used by the UTRAN for protecting signalling messages exchanged between the mobile and the UTRAN.

IMEI *International Mobile Station Equipment Identity*. Uniquely identifies the mobile terminal.

IMSI *International Mobile Station Identity*. Uniquely identifies the GSM subscriber.

Integrity checks Mechanisms intended to guarantee to the receiver of a signalling message that the content of the message has not been altered or falsified during transmission.

IS-95 Second generation CDMA standard, used mainly on the American continent and in Asia.

IS-136 Second generation TDMA standard, used mainly in North America.

LA *Location Area*. Geographic zone used by the core network, in particular by the MSC/VLR during paging, for locating subscribers in idle mode.

Logical channel To each type of information (paging messages, system information, etc.) carried over the radio interface, there corresponds a separate logical channel.

MAC *Medium Access Control*. UTRAN radio interface layer responsible for multiplexing data over the radio channels.

MAP *Mobile Application Part.* The MAP describes the signalling and services (call set-up, mobility, subscriber management, etc.) supported by the GSM and UMTS core network standards.

MIB *Master Information Block.* Master block telling the mobile how the system information (or SIB) is broadcast on the BCCH.

MM *Mobility Management.* GSM and UMTS network protocol responsible for the mobility (in the sense of the core network) and security aspects of circuit mode calls.

MSC *Mobile Service Switching Center.* Switch used in the GSM and UMTS core networks for circuit mode services.

MVNO *Mobile Virtual Network Operator.* An operator providing telecommunication services who only owns a part of the network infrastructure.

NAS *Non-Access Stratum.* UMTS network layer containing the functions that are independent of the access layer (AS).

NBAP *Node B Application Part.* Signalling protocol used on the lub interface (i.e. between the RNCs and the Node Bs).

Node B UTRAN entity responsible mainly for transmission over the radio interface.

NSS *Network Sub-System.* GSM core network.

OVSF *Orthogonal Variable Spreading Factor.* A family of codes used on the UTRAN radio interface.

Paging Procedure involving broadcasting a message to a subset of cells when the network wants to set up a call with a mobile user.

PCCH *Paging Control Channel.* Logical control channel used for sending paging messages.

PCH *Paging Channel.* A unidirectional shared transport channel.

PDC *Personal Digital Cellular.* Cellular second generation radio-communication system used in Japan.

PDCP *Packet Data Convergence Protocol.* UTRAN radio interface layer responsible for compressing and decompressing the headers of packets transmitted.

PDP context A context containing the information relating to a packet-switched call for a given user. Several PDP contexts can be simultaneously active for the same user.

PDSCH *Physical Downlink Shared Channel.*

PHS *Personal Handyphone System.* Cellular second generation radio-communication system used in Japan.

Physical channel A resource on the radio interface used for the transmission of information.

PLMN *Public Land Mobile Network.* UMTS or GSM cellular radio-communication network run by an operator in a given geographical area.

PRACH *Physical Random Access Channel.* Uplink physical channel used by mobiles for access to the network and for transmitting low speed user traffic in the RRC CELL_FACH state.

PS domain Subset of the UMTS core network responsible for packet-switched applications.

PSTN *Public Switched Telephone Network.*

QPSK *Quadrature Phase Shift Keying*. Phase modulation technique used on the UTRAN radio interface.

RA *Routing Area*. Geographic zone used by the core network, in particular, the SGSN, for locating subscribers in idle mode.

RAB *Radio Access Bearer*. Data transmission service provided by the access stratum to the non-access stratum.

RACH *Random Access Channel*. Shared unidirectional uplink channel associated with the PRACH.

RANAP *Radio Access Network Application Part*. Signalling protocol used on the Iu interface (i.e. between the RNCs, the SGSN and the MSC).

Relocation Procedure for changing the UTRAN anchor point (the SRNC) of a mobile/network connection.

Reselection A procedure that allows a mobile that is switched on to change its reference cell. This procedure is employed in idle mode (no mobile/network connection) and in certain connected mode states (RRC URA_PCH, CELL_PCH, and possibly CELL_FACH).

RLC *Radio Link Control*. UTRAN radio interface layer responsible for reliable data transmission.

RNC *Radio Network Controller*. UTRAN entity responsible mainly for routing calls in the access network and for control of radio resources.

RNSAP *Radio Network Subsystem Application Part*. Signalling protocol used on the Iur interface (i.e. between the UTRAN RNCs).

RRC *Radio Resource Control*. UTRAN radio interface layer responsible for controlling resources used.

RTT *Radio Transmission Technology*. Generic term designating the access technique on the radio interface of cellular networks.

Selection Procedure used when the mobile is switched on, in order to identify a PLMN cell suitable for providing communication services to the user.

SGSN *Serving GPRS Support Node*. GSM/GPRS and GSM core network equipment containing data about subscribers in idle mode (such as the current location area). The SGSN is the packet mode equivalent of the MSC/VLR.

SIB *System Information Block*. Broadcast by the UTRAN on the BCCH.

SIM *Subscriber Identification Module*. A smart card used in GSM mobile terminals, which contains information about the subscriber.

SM *Session Management*. GSM/GPRS and UMTS network protocol responsible for handling packet-mode calls.

SMS *Short Message Service*.

TCP *Transmission Control Protocol*. Transport protocol used on the Internet.

TDD *Time Division Duplex*. Separation of uplink and downlink channels by using different transmission times.

TDMA *Time Division Multiple Access*.

TFS *Transport Format Set*. Set of transport formats possible for a given RAB.

Transport channel A transport channel is the means by which information is transported over the radio interface and is characterised by such parameters as block size, type of channel coding, etc.

TRAU *Transcoder and Rate Adaptation Unit*. Component of GSM and UMTS networks responsible for coding speech ready for transmission over the radio interface.

UEA *UMTS Encryption Algorithm*. Data encryption algorithm used in the UTRAN.

UIA *UMTS Integrity Algorithm*. Algorithm to guarantee the integrity of signalling messages transmitted between mobiles and the UTRAN.

UMSC *UMTS MSC*. UMTS core network component, bringing together the functions of the MSC/VLR and the SGSN.

UMTS *Universal Mobile Telecommunication System*.

URA *UTRAN Registration Area*. Set of cells that allows the UTRAN to locate a subscriber in the RRC URA_PCH connection state.

USIM *Universal Subscriber Identity Module*. UMTS equivalent of the SIM card used in GSM terminals.

UTRAN *Universal Terrestrial Radio Access Network*. The UMTS access network.

VLR *Visitor Location Register*. GSM and UMTS core network database containing data about idle mode subscribers (the current location area, for example).

Index

F

G

H

I

K

L

M

N

O

P

Q

R

T

V

W

X