UMTS Networks

Architecture, Mobility and Services

Heikki Kaaranen
Aqua Records Oy, Finland

Ari Ahtiainen
Nokia Research Center, Finland

Lauri Laitinen
Nokia Research Center, Finland

Siamäk Naghian
Nokia Networks Oy, Finland

Valtteri Niemi
Nokia Research Center, Finland

JOHN WILEY & SONS, LTD

Chichester • New York • Weinheim • Brisbane • Singapore • Toronto

National 01243 779777
International (+44) 1243 779777
e-mail (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on http://www.wiley.co.uk or http://www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons, Inc., 605 Third Avenue,
New York, NY 10158-0012, USA

WILEY-VCH Verlag GmbH
Pappelallee 3, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton,
Queensland 4064, Australia

John Wiley & Sons (Canada) Ltd, 22 Worcester Road
Rexdale, Ontario, M9W 1L1, Canada

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01,
Jin Xing Distripark, Singapore 129809

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Image on front cover was supplied by Xplorer (www.xplorer.co.uk).

ISBN 0471 48654 X

Typeset in Times by Deerpark Publishing Services Ltd, Shannon.
Printed and bound in Great Britain by Biddles Ltd, Guildford and King's Lynn.
This book is printed on acid-free paper responsibly manufactured from sustainable forestry, in which at least two trees are planted for each one used for paper production.

Contents

UMTS system. Chapter 1 introduces the UMTS technical and service architecture and key system concepts. Chapter 2 is an illustrated story about mobile network evolution from second generation GSM into the first UMTS release and beyond that towards full-IP mobility networks.

The second part consists of Chapters 3–8, which examine the radio access and core network in more detail, explaining the functions and services provided to the end users. Chapter 3 on UMTS radio communications provides the fundamentals of cellular radio which are necessary for understanding the WCDMA radio access system as part of the UMTS network. Chapters 4 and 5 present the functional split and system management aspects distributed among the UMTS network elements in the radio access and core network parts. Chapter 6 provides an overview on the UMTS user equipment focusing on those aspects, which are most visible to the rest of the UMTS network. In Chapter 7 the UMTS network is examined as a network for services. The UMTS service capabilities (WAP, CAMEL, location based services, etc.) are first introduced as a value-added service platform and then application examples with different Quality of Service (QoS) characteristics are discussed. The advanced security solutions of the UMTS network are then discussed in Chapter 8.

The remaining Chapters 9 and 10 form the third part of the book. In these chapters we take a protocol-oriented view describing the system-wide interworking between the different architectural elements. Chapter 9 first elaborates the basic UMTS protocol architecture and then introduces the individual system protocols one by one. Chapter 10 returns to the network-wide view by showing selected examples of system procedures, which describe how the transactions are carried out across the UMTS network interfaces under the co-ordination of the system protocols.

At the footstep of the third generation mobile communications the success of the UMTS will come from the thousands of leading system and software engineers, content providers, application developers, system integrators and network operators. We hope this book will help all of them to reach their targets and let them enjoy and benefit from the UMTS networking environment.

This book represents the views and opinions of the authors and does not necessarily represent the views of their employers.

Acknowledgements

During the "UMTS Book project" the authoring team members have had a pleasure to follow the exciting finalisation of the UMTS system specifications. Among the colleagues both from Nokia and outside, those listed below have provided valuable input and comments on the various aspects of the book: Seppo Alanara, Mika Forssell, Harri Holma, Kaisu Iisakkila, Tatjana Issayeva, Pekka Korja, Jan Käll, Juho Laatu, John Loughney, Atte Länsisalmi, Anna Markkanen, Tomi Mikkonen, Juha Mikola, Ahti Muhonen, Mikko Puuskari, Mikko J. Rinne, Ville Ruutu, Juha Sipilä, Janne Tervonen, Mikko Tirronen, Ari Tourunen and Jukka Vialén.

This inspiring working environment and close contacts with the R&D and standardisation programs within Nokia were made possible by the following managers in those programs: Kari Aaltonen, Heikki Ahava, Tapio Harila, Jari Lehmusvuori, Juhani Kuusi, Yrjö Neuvo, Pertti Paski, Tuula-Mari Rautala, Tuomo Sipilä, Jukka Soikkeli, Jari Vainikka and Asko Vilavaara.

The publishing team lead by Mark Hammond at John Wiley & Sons Ltd. has participated in the production of this book by giving us excellent support. Their hard working spirit made it possible to keep the demanding schedule in the publication process.

After all, this is a book about UMTS networks, which are based on the joint design and engineering effort of numerous colleagues of ours, who have brought their expertise together to make it happen. Without being able to list all the experts from the early 1990s until the recent meetings of the 3GPP organisation, we like to thank all of them for their dedicated work.

Finally we like to express loving thanks to all the members of our families for the patience and support shown during the long days and late nights of the book writing effort. Among them, Mrs Satu Kangasjärvelä-Kaaranen deserves special thanks, since her help in word processing and graphical design of numerous figures was invaluable in putting together the manuscript.

The authors welcome any comments and suggestions for improvements or changes that could be implemented in possible new editions of this book. The e-mail address for gathering such input is umtsnetworks@pcuf.fi.

Helsinki, Finland.
The authors of *UMTS Networks*.

W

W	Watt
WAP	Wireless Application Protocol
WARC	World Administrative Radio Conference
WCDMA	Wideband Code Division Multiple Access
WLAN	Wireless Local Area Network
WML	Wireless Markup Language
WTLS	Wireless Transport Layer Security
WWW	World Wide Web

X

X.25	An ITU-T Protocol for Packet Switched Networks
X.509	Internet X.509 Public Key Infrastructure
XMAC	Expected Message Authentication Code
XRES	Expected user Response

1

Introduction

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

Because the need for mobile communication increased, also the need for a more global mobile communication system increased. The international specification bodies started to

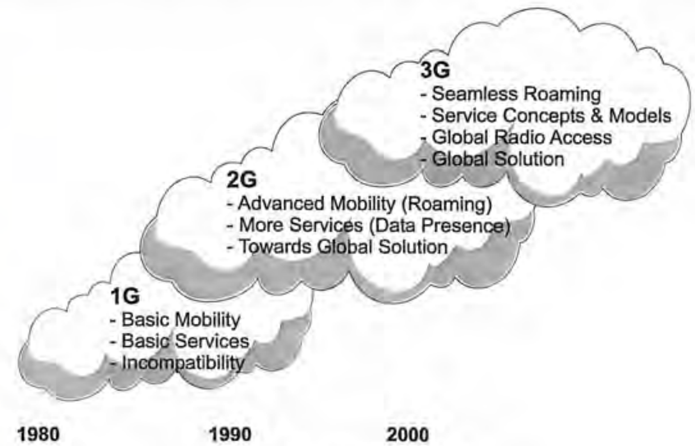


Figure 1.1. Cellular generations

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

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

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

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

The 3G system is therefore already in evolution through new phases and recently the discussion on 4G has already started. Right now it may be too early to predict where the 3G evolution ends and 4G really starts. Rather this future development can be thought of as an ongoing development chain where 3G will continue to introduce new ways to handle and combine all kinds of data and mobility. 4G will then emerge as a more sophisticated system concept bringing still more added value to the end-users.

1.1 Specification Process for 3G

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

A common, global mobile communication system naturally creates a lot of political desires. In the case of 3G this can be seen even in the naming policy of the system. The most neutral term is third generation, 3G. In different parts of the world different issues are emphasised and thus the global term 3G has regional synonyms. In Europe 3G has become UMTS (Universal Mobile Telecommunication System), following the ETSI perspective. In Japan and the US the 3G system often carries the name IMT-2000 (International Mobile Telephony 2000). This name comes from the International Telecommunication Union (ITU) development project. In the US the CDMA2000 is also one aspect of 3G cellular systems and it represents the evolution from the IS-95 system. In this book, we will describe the UMTS system as it has been specified by the world-wide 3G Partnership Project (3GPP).

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

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

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

ITU had a dream to specify at least one common global radio interface technology. This kind of harmonisation work was done under the name Future Public Land Mobile Telephony

System (FPLMTS) and later IMT-2000. Due to many parallel activities in regional standardisation bodies this effort turned into promotion of common architectural principles among the family of IMT-2000 systems.

Europe and Japan also had different short-term targets for 3G system development. In Europe a need for commercial mobile data services with guaranteed quality, for example mobile video services, was widely recognised after the early experiences from narrowband GSM data applications. Meanwhile in the densely populated Far East there was an urgent demand for additional radio frequencies for speech services. The frequency bands identified by ITU in 1992 for the future 3G system called IMT-2000 became the most obvious solution to this issue. In early 1998 a major push forward was achieved when ETSI TC-SMG decided to select WCDMA as its UMTS radio technology. This was also supported by the largest Japanese operator NTT DoCoMo. The core network technology was at the same time agreed to be developed on the basis of the GSM core network technology. During 1998 the European ETSI and the Japanese standardisation bodies (TTC and ARIB) agreed to make a common UMTS standard. After this agreement, the 3GPP organization was established (Figure 1.2) and the determined UMTS standardisation was started worldwide.

From the UMTS point of view, the 3GPP organisation is a kind of "umbrella" aiming to

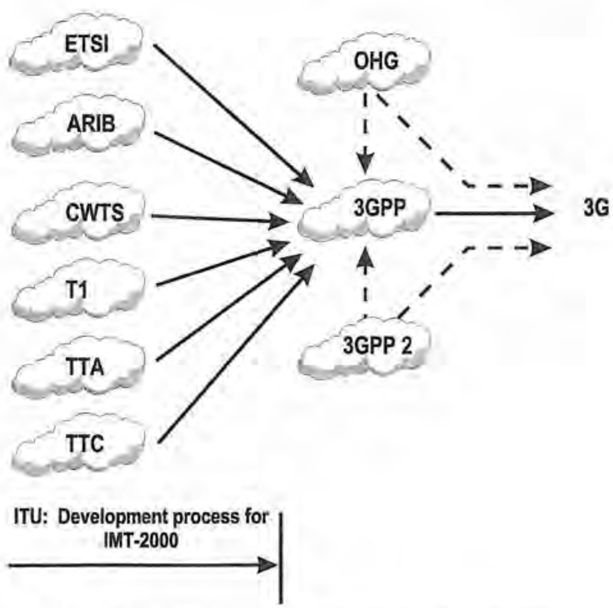


Figure 1.2. UMTS standardisation procedure – principle diagram

form compromised standards by taking into account political, industrial and commercial pressures coming from the local specification bodies:

- ETSI (European Telecommunication Standard Institute)/Europe
- ARIB (Association of Radio Industries and Business)/Japan
- CWTS (China Wireless Telecommunication Standard group)/China
- T1 (Standardisation Committee T1 – Telecommunications)/US
- Telecommunication Technology Association (TTA)/Korea
- Telecommunications Technology Committee (TTC)/Japan

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

To ensure that the American viewpoint will be taken into account a separate Third Generation Partnership Project Number 2 (3GPP-2) was founded and this organisation performs specification work from the IS-95 radio technology basis. The common goal for 3GPP, OHG and 3GPP-2 is to create specifications according to which a global cellular system having wideband radio access could be implemented.

To summarise, there was three different approaches towards the global cellular system, 3G. These approaches and their building blocks are, on a rough level, presented in Table 1.1.

Table 1.1 3G variants and their building blocks

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

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

The 3GPP originally decided to prepare specifications on a yearly basis, the first specification release being Release 99 (3GPP R99). This first specification set has a relatively strong "GSM presence". From the UMTS point of view the GSM presence is very important; first, the UMTS network must be backward compatible with the existing GSM networks and second, the GSM and UMTS networks must be able to inter-operate together. The next release was known as 3GPP R00 but because of the multiplicity of changes proposed, the specification activities were scheduled into two specification releases 3GPP R4 and 3GPP R5. Also – to be consistent – the 3GPP R99 is sometimes called 3GPP R3. The 3GPP R4 defines major changes in UMTS core network circuit switched side and those are related to the separation of user data flows and their control mechanisms. Another big item in 3GPP R4 is to introduce mechanisms and arrangements for multimedia. An overview of 3GPP R4 is given in Chapter 5. The 3GPP R5 aims to introduce a UMTS network where the transport network utilises IP networking as much as possible. Therefore this goal is called the "All IP" network. IP and overlying protocols will be used in network control, too and also the user data flows are expected to be mainly IP based. In other words, the mobile network implemented according to the 3GPP R5 specifications will be an end-to-end packet switched cellular network using IP as the transport protocol instead of SS7 (System Signalling 7), which holds the major position in existing circuit switched networks. Naturally the IP-based network should still support circuit switched services, too. The 3GPP R4/R5 will also start to utilise the possibility for new radio access techniques. In 3GPP R99 the basis for the UMTS Terrestrial Access Network (UTRAN) is WCDMA radio access. In 3GPP R4/5 another radio access technology derived from GSM with Enhanced Data for GSM Evolution (EDGE) will be specified to create GSM/EDGE Radio Access Network (GERAN) as an alternative to building a UMTS mobile network.

1.2 Introduction to 3G Network Architecture

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

- Conceptual network model
- Structural network architecture
- Resource management architecture
- UMTS service and bearer architecture

1.2.1 Conceptual Network Model

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

From the protocol structure and their responsibility point of view, the 3G network can be divided into two strata: access stratum and non-access stratum. *Stratum* refers to the way of grouping protocols related to one aspect of the services provided by one or several domains, see 3GPP specification TR 21.905. Thus, the access stratum contains the protocols handling activities between the User Equipment (UE) and access network. The non-access stratum contains protocols handling activities between the UE and Core Network (CS/PS domain), respectively. For further information about strata and protocols refer to Chapter 9.

The part of Figure 1.3 called "Home Network" maintains static subscription and security information. The serving network is the part of the core network + domain, which provides the core network functions locally to the user. The transit network is the core network part located on the communication path between the serving network and the remote party. If, for a given call, the remote party is located inside the same network as the originating UE, then no particular instance of the transit network is needed.

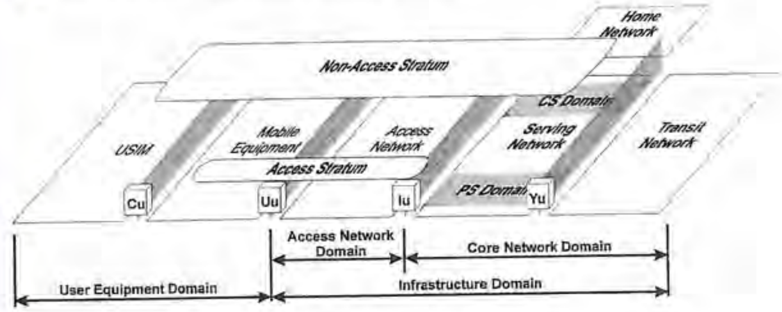


Figure 1.3. UMTS architecture - conceptual model

1.2.2 Structural Network Architecture

In this book we mainly present the issues from the network structural architecture perspective. This perspective is presented in Figure 1.4. In UMTS the GSM technology has a remarkable role as background and actually UMTS aims to reuse everything, which is reasonable. For example, some procedures used within the non-access stratum are, in principle, reused from GSM but naturally with required modifications.

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

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

The UTRAN is divided into Radio Network Subsystems (RNS). One RNS consists of a set of radio elements and their corresponding controlling element. In UTRAN the radio element is Node B, referred to as Base Station (BS) in the rest of this book, and the controlling element is Radio Network Controller (RNC). The RNSs are connected to each other over access network-internal interface Iur. This structure and its advantages are explained in more detail in Chapter 4.

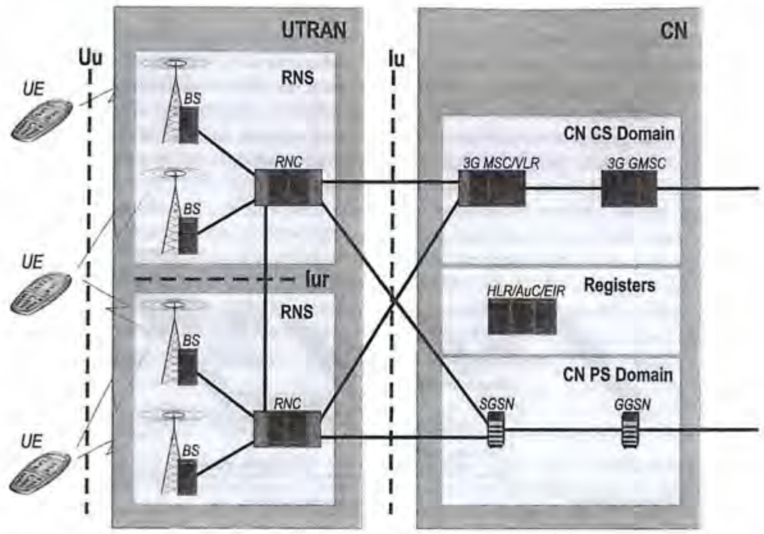


Figure 1.4. UMTS network architecture – network elements and their connections for user data transfer

The term Core Network (CN) covers all the network elements needed for switching and subscriber control. In early phases of UMTS part of these elements are directly inherited from GSM and they are modified for UMTS purposes. Later on, when transport technology changes, the core network internal structure will remarkably change, too. The term CN covers Circuit Switched (CS) and Packet Switched (PS) domains defined in Figure 1.3. The configuration alternatives and elements of the UMTS core network are discussed in Chapter 5.

The part of the Figure 1.4 called "Registers" is the same as Home Network in the preceding 3G network conceptual model. This part of the network maintains static subscription and security information. Registers are discussed in more detail in Chapter 5.

The major open interfaces of UMTS are also presented in Figure 1.4. Between the UE and UTRAN the open interface is Uu, which in UMTS is physically realised with WCDMA technology. For further details about WCDMA, refer to Chapter 3. The other major open interface is Iu located between the UTRAN and CN.

The RNSs are separated from each other with an open interface Iur. Iur is a remarkable difference when compared to GSM; it brings completely new abilities for the system to utilise so called macro diversity and also efficient radio resource management and mobility mechanisms. When the Iur interface is implemented in the network, the UE may attach to the network through several RNCs and each of those maintains a certain logical role during the radio connection. These roles are Serving RNC (SRNC), Drifting RNC (DRNC) and Controlling RNC (CRNC). The Controlling RNC has the overall control of the logical resources of its UTRAN access points, being mainly BSs. A SRNC is a role an RNC can take with respect to a

specific connection between the UE and UTRAN. There is one SRNC for each UE that has a radio connection to UTRAN. The SRNC is in charge of the radio connection between the UE and the UTRAN. It also maintains the Iu interface to the core network, which is the main characteristic of the SRNC. A DRNC is a logical role used when radio resources of the connection between the UTRAN and the UE need to use cell(s) controlled by another but the SRNC itself. The UTRAN related issues in general are discussed in Chapter 4.

In addition to the CS and PS domains presented in Figure 1.4 the network may contain other domains. One example of these is the broadcast messaging domain, which is responsible for multicast messaging control. In this book we however concentrate on the UMTS network like that presented in Figure 1.4.

1.2.3 Resource Management Architecture

The network element–centric architecture described above results from functional decomposition and the split of responsibilities between major domains and finally between network elements. Figure 1.5 illustrates this split of major functionalities, which are

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

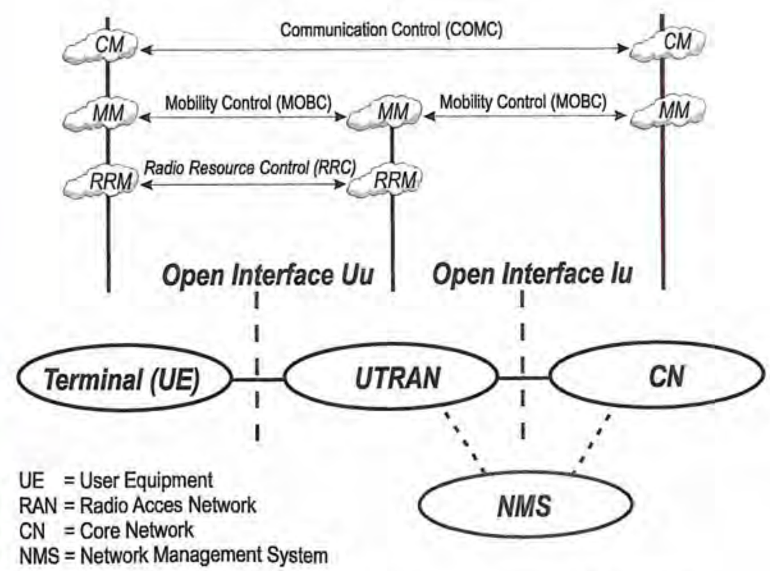


Figure 1.5. UMTS network architecture – management tasks and control duties

CM covers all of the functions and procedures related to the management of user connections. CM is divided into several sub-areas such as call handling for circuit switched connections, session management for packet switched connections, as well as handling of supplementary services and short message services. MM covers all of the functions and procedures needed for mobility and security, for example, connection security procedures and location update procedures. Most of the MM procedures occur within the core network and its elements but in the 3G part of the MM functions are performed also in UTRAN for packet switched connections. CM and MM principles are discussed in Chapter 5.

RRM is a collection of algorithms UTRAN uses for management of radio resources. These algorithms handle, for instance, power control for the radio connections, different types of handovers, system load and admission control. The RRM is an integral part of UTRAN and basic RRM is discussed more closely in Chapter 4. Some system-wide procedure examples about CM, MM and RRM functioning are given in Chapter 10.

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

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

COMC maintains mechanisms like call control, and packet session control. MOBC maintains mechanisms, which cover, for example, execution control for location updates and security. The radio resources are completely handled within UTRAN and UE. The control duty called RRC takes care of, e.g. radio link establishment and maintenance between the UTRAN and UE. These collections of control duties are then further refined into a set of well-specified control protocols. For more detailed information about protocols, refer to Chapter 9.

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

1.2.4 UMTS Service and Bearer Architecture

Both in 1G and 2G networks the technology implementation and its details were issues as such. In 3G the technology is important but the common opinion is that in 3G the network is more a "service network" than a plain cellular network. In other words, 1G and 2G were technologically limited networks allowing the end-users to use a limited set of technology-specific services. In 3G the technology should not be the limitation. This kind of approach is partially commercial and, in a way, gets the operator thinking. If this kind of service viewpoint is discussed, a 3G network can be modelled as shown in Figure 1.6.

Because a 3G network will be a very complex infrastructure as a whole, there are two

Figure 1.6. UMTS network architecture – service model

issues, which are present everywhere in the network and should be handled carefully. These are network management and security. From the service architecture standpoint, a 3G network and its building elements can be divided into four different layers. The first layer being the basis of all is physical transmission. The nodes using physical transmission form their own layer, network elements. The third layer contains elements and functionalities forming the environment where the services for the end-users are created. Since the sophisticated services require contents, the content provision is separated into its own layer on top of the stack. It is often stated that the lower layers indicated in Figure 1.6 require the greatest investments and on the higher layers the investments are not playing the major role. Rather it can be said that every layer in this model generates both costs and revenue but those are different in nature. The two following principles can be applied to this four-layer stack:

- Rule A: the lower location the layer has, the bigger is the investment in network elements. In other words, transmission and network elements form the biggest cost in the 3G network. Revenue generation depends on the network configuration and coverage.
- Rule B: The higher location the layer has, the bigger is the investment in people and ideas. When moving to the higher layers, the pure technology is not an issue as such; more people having good ideas and an understanding of human behaviour are the core business principles here.

As a conclusion, a profitable and successful 3G network is a successful compromise and combination of those layers presented in Figure 1.6. In other words, technology forms a platform where good ideas have the possibility to turn into successful services for the end-users.

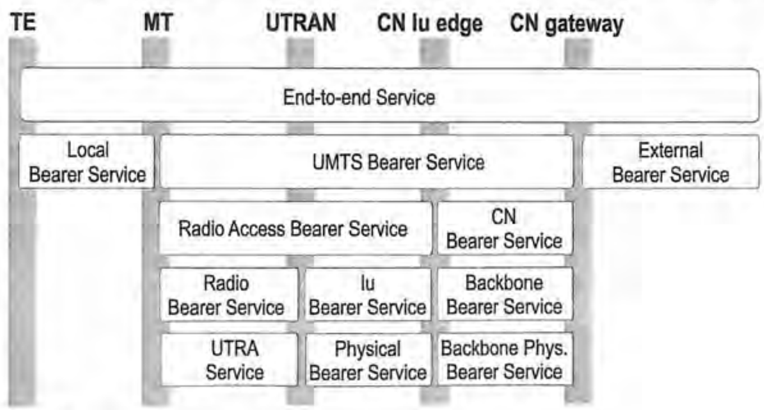
In traditional 2G networks the role of the network management is relatively limited and concerns normally the network elements and their technical behaviour. In 3G the idea of separating services and platforms and network widens the scope of network management; every service used requires involvement from every layer of this model and in order to ensure the correct functionality of the service the whole stack should be controlled.

The access level and network-internal security in UMTS is ensured by utilising very sophisticated mechanisms and effective algorithms. In this respect UMTS networks are more secure than GSM networks. The open business model supporting third party involvement brings in the aspect of end-to-end security. This is, however, not in the scope of 3GPP specifications though the potential risks involved are recognised. Security in the UMTS environment is discussed more closely in Chapter 8.

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

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

Within the UMTS network the QoS handling is different in UTRAN and CN. From the CN point of view the UTRAN creates an "illusion" of a fixed bearer providing adequate QoS for the end-user service. This "illusion" is called radio access bearer service. Within the CN an own type of bearer service called CN bearer service is used. This division between radio



Legend:
TE = Terminal Equipment
MT = Mobile Termination

Figure 1.7 Bearer architecture in UMTS

access bearer and CN bearer service is required since the QoS must be guaranteed in very different environments and both of these environments require their own mechanisms and protocols. For instance, CN bearer service is quite constant in nature since the backbone bearer service providing physical connections is also stable. Within UTRAN the radio access bearer experiences more changes as a function of time and UE's movement and this sets different challenges for QoS. This division pursues also the main architectural principle of the UMTS network, that is, the independence of the entire network infrastructure from the radio access technology.

The structure presented in Figure 1.7 is a network architecture model from the bearer and QoS point of view. Since the QoS is one of the most important issues in UMTS, the QoS and bearer concepts are handled throughout this book.

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

3

Basics of UMTS Radio Communication

The purpose of this chapter is to introduce briefly two fundamental radio communication issues to provide the reader, first, the reasons behind those restrictions and also opportunities when considering the basic architecture of any radio communication system. The second issue is related to how the radio communication is specifically handled in UMTS. This is addressed by describing the essentials of the third generation (3G) radio path and more specifically WCDMA-FDD, providing substantial mechanisms and terminology. More insight into the 3G radio path characteristics is given in Holma and Toskala (2001).

3.1. Radio Communication Fundamentals

Communication has always been the essential part of every kind of society and especially human society. Because social evolution has been an inherent characteristic of collective life resulting in more sophisticated interaction relationships, more advanced means have always been needed to meet the communication demands between members of society. During the history of human life, many techniques have been developed for the purpose of communication. Among them, radio communication has been, and will be one of the most important techniques that man has ever used for communicating.

The usage of radio communication has been realised since Hertz experimentally showed the relationship between light and electricity in 1887 after Maxwell had illustrated the principal equations of electromagnetic fields in 1864. Then, Marconi used the "Hertzian" wave to communicate, inventing wireless telegraphy in 1896. Exploiting the radio wave for transmitting information is the basic part of every kind of radio communication.

The fundamental principle of radio communication is that it utilises radio waves as a transmission medium. As natural phenomena, radio waves are a consequence of electromagnetic fields. Under certain circumstances, time dependent electromagnetic fields produce waves that radiate from the source to the environment. This source can be for example a transmitter like a base station or mobile handset. Because radio waves in general are based on electromagnetic fields, their characteristics are strictly dependent on the environment where the waves propagate. As a consequence, a radio wave based radio communication system is vulnerable to the environmental factors, for instance, mountains, hills, huge reflectors like buildings, the atmosphere, and so on.

Every communication system consists of at least two elements, which are the transmitter and the receiver. As it is the case in mobile systems, these two elements can be integrated in one device (transceiver) so that it is capable of operating both as a transmitter and receiver device. An example of such a device is the base station and mobile handset in any advanced public mobile system. Figure 3.1 illustrates the simplest radio communication system, consisting of one base station and one mobile handset. Suppose that the base station acts as a transmitter source for a specific time with certain environmental circumstances. Then the radio signal propagates from base station to the mobile station and with the speed of light. The received signal strength at the handset depends on the distance from the base station, the wavelength, and the communication environment.

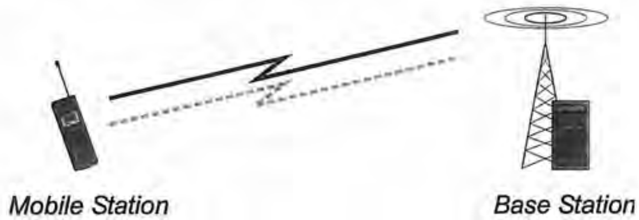


Figure 3.1. Essential elements of a radio communication system

A radio wave propagation mechanism closely depends on the wavelengths or frequencies. In addition to that, any man-made or natural obstacle like high buildings, terrain, weather condition, etc. affects the way and time of signal propagation between the transmitter and receiver. Similarly, system parameters, for example antenna height and beam direction have naturally their own effect on the propagation distance, mode, and delay. The nature of radio communications inherently brings about some thorny problems. The main problems that every radio communication faces are as follows:

- Multipath propagation phenomena
- Fading phenomena
- Radio resource scarcity

Multipath propagation is also considered, by many means, as an advantage of the radio communication because it enables the radio receiver to hear the base station even without signal Line Of Sight (LOS). Despite that, it brings complexity to the system by setting specific requirements and constraints for it. In order to understand the nature of a radio communication system those characteristics should be well understood. Therefore, we explain them here in more detail.

The factors that effect radio propagation are extremely dynamic, sophisticated, and diverse. In spite of that and in order to model the propagation phenomena, the mechanisms can be classified into reflection, diffraction, and scattering phenomena (Figure 3.2). In the radio network environment these propagation mechanisms lead to multipath propagation, which causes fluctuations in the received signal's characteristics, including amplitude, phase, and angle of arrival, giving rise to multipath fading. Reflection is the consequence of collision

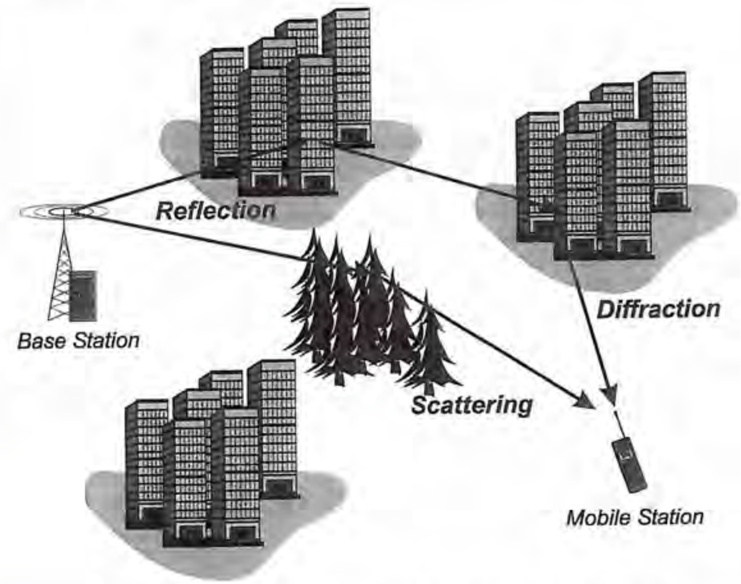


Figure 3.2. Propagation mechanisms: reflection, scattering and diffraction

of the electromagnetic wave with an obstruction whose dimensions are very large in comparison with the wavelength of the radio wave. The result of this phenomenon is reflected radio waves, which can be captured constructively at the receiver, e.g. mobile or base station. Diffraction, also called shadowing, in turn, is the consequence of collision of the radio wave with a obstruction which is impossible to penetrate. Scattering on the other hand is the consequence of the collision of the radio wave with obstructions whose dimensions are almost equal to or less than the wavelength of the radio wave. These phenomena together explain how radio waves can travel in a radio network environment without an LOS path.

From the receiver's perspective and depending on the existing preconditions of any above-mentioned propagation mechanisms, the received signal power is affected randomly by each or combination of these mechanisms. In addition to those propagation mechanisms, in mobile system the concepts of mobility, indoor coverage, outdoor coverage, and hierarchical network structure raise some specific aspects to the propagation environment, which makes the situation more sophisticated.

There are mainly two different ways to describe the effect of the propagation mechanism on the signal strength of the radio channel, including link budget and time dispersion. The basic idea behind the link budget is to determine the expected signal level at a particular distance or location from a transmitter like base station or mobile station. By modelling the link budget essential parameters of the radio network such as transmitter power requirements, coverage area, and battery life can be defined.

Link budget calculation, can be carried out by estimating the signal path-loss. Path-loss estimation can be done based on the free space model, defining that in an idealised free space model the attenuation of signal strength between the base station and mobile station likely behaves according to an inverse-square law. Due to radio environment differences in advanced mobile systems, it is almost impossible to consider all parameters affecting the radio channel modelling and system designing. Therefore, some general models, for the most usual cases have been developed (Okumura et al.). In order to consider the entire effect of radio channel fading, however, link budget alone cannot be adequate. In addition to that the effect of multipath propagation in terms of time dispersion should also be considered. This can be done by estimating the different propagation delays related to the replicas of the transmitted signal which reaches the receiver.

A typical fading process is illustrated in Figure 3.3. As shown, any fading process has the curves of two different simple shapes, that is, the downward deep fades, referring to the deterioration of the signal strength and the upraise curves, causing the undesired interference. Therefore, a combination of these simple curves can approximate the envelope of any fading process and the control actions for the fading process are separately determined to compensate for the round deviation of the signal level around the desirable average.

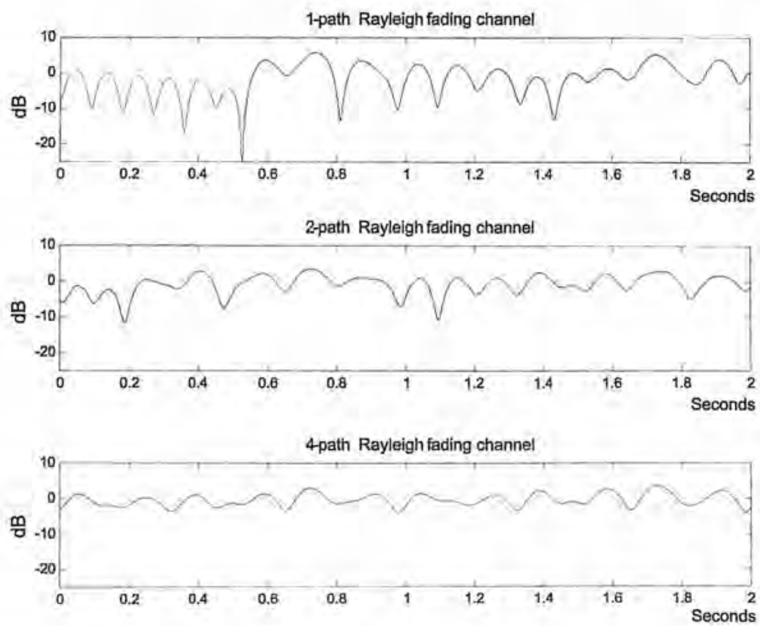


Figure 3.3. A typical fading power signal

Figure 3.4 illustrates the main fading types, which may arise in any radio network environment in one way or another depending on the radio network environment. The main classes of the radio channel fading include large-scale and small-scale fading. The former represents the average signal power attenuation or path loss due to the mobile's motion over areas between the base station and the mobile station. Small-case fading, on the other hand, is the result of rapid variations in signal amplitude and phase that can be experienced between the base station and mobile station. Small-scale fading is also called Rayleigh fading or Rician fading depending on the NLOS or LOS characteristic of the reflective paths. Small-scale fading can be divided into frequency selective fading, flat fading, fast fading, and slow fading.

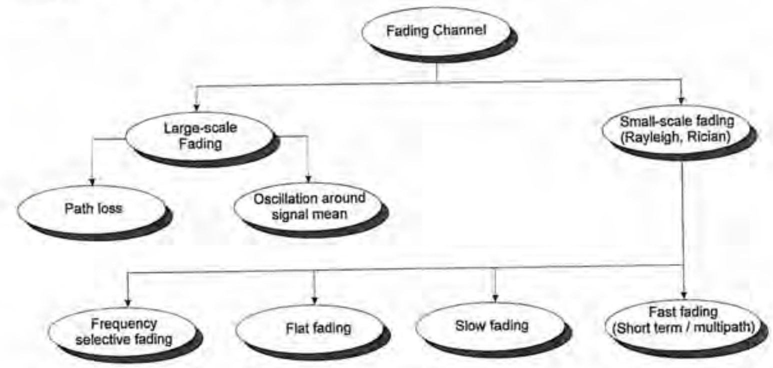


Figure 3.4. Radio channel fading classification

The radio channel experiences signal fading that most likely consists of the Rayleigh faded component without LOS, making the channel more severe to cope with. As a result, a signal arrives at a mobile station from many directions with different delays, causing significant phase differences in signals travelling on different branches. Moreover, the behaviour of the radio channel fading is strictly dependent on the physical position and motion of the mobile station. A small change in position or motion of the mobile station may result in a different phase for each signal branch. Thus the motion of the mobile station through the radio network may result in a quite rapid fading phenomenon, which can be extremely disruptive to the radio signals, requesting considerable requirements from the radio network optimisation.

If we assume that a radio signal is transmitted from the base station to the mobile station with constant transmitted signal strength, then, the effect of fading can be seen as a function of distance of the strength of the base station transmitted signal as received at a mobile station. Therefore, as distance increases between the mobile station and base station, the received signal strength decreases at the mobile station. The average path loss – which basically follows log-normal distribution of average path loss – is the same for both directions.

Multiple path signal propagation causes rapid fading as well. Indeed, this type of fading is more difficult to compensate for because the signal arrives from these multiple paths in random phases and amplitude, resulting in Rayleigh fading characteristics. If different bandwidths are used in uplink and downlink in cellular systems then the Rayleigh fading is typically independent for downlink and uplink, meaning that when the uplink channel is fading, the downlink channel is not necessarily fading at the same time and vice versa.

In order for the mobile to adjust its transmitter power to correspond to the link path signal strength of its home cell base station it should take the inverse pace as the signal the base station experiences. Therefore, the mobile has to power up its signal in two control steps, that is, firstly it should compensate the log-normal fading and secondly fast fading. Sudden increase in transmitter power, however, may have drawbacks for the overall system performance in terms of capacity and service quality. As mobility is an inherent part of the mobile system, it causes more complexity to the radio channel fading. The duration of the fades from the median signal level is dependent on mobile station speed and the operating frequency in such a way that as the mobile station accelerates the fading becomes more severe.

In addition to fading and multipath phenomena, interference is a thorny problem for every radio communication system. The basic reason behind interference is that there are many simultaneous radio accesses to the base station. This is the case especially when the radio system uses a common share bandwidth. Moreover, many man-made noises may generate additional interferences to the radio system. However, the most dominant interference modes are basically caused by the common use of radio resources simultaneously. Minimising the undesirable impact of fading, interference and radio resource scarcity is considerably dependent upon the radio network planning approach, the utilised radio access techniques, and those algorithms used for controlling the radio resources. Therefore, before describing further details of the interference aspects of a radio system, we proceed next by introducing the concept of a cellular system as a basic solution for capacity limitation of radio systems due to path-loss, fading and interference.

In order to solve the previously mentioned problems many solutions have been developed especially during last decades since radio communications have been used for public communications. An increase in the number of radio communication users together with simultaneous demand for large area network coverage and diverse services, however, set extremely tight requirements for radio communication systems. As a result, many advanced solutions have been utilised by public radio communication systems, e.g. cellular concepts, advanced radio resource allocation techniques, modulation techniques, advanced antenna techniques, and so on.

3.2. Cellular Radio Communication Principles

The simple radio communication system illustrated earlier in this chapter is unable to provide access service to the large number of end-users, that is, it faces capacity limitation. Let's outline the basic problems that the inherent characteristics of radio communications described earlier cause for such a simple system.

Firstly, the public radio communications should offer duplex communication for providing simultaneous two way communication. Therefore, it is not practical to offer such services by, for instance broadcasting the information. In addition to that, the signal strength at the [illegible] between the transmitter and receiver increases, resulting

in unacceptable QoS in the area far away from the transmitter. Secondly, every transmitter is capable of offering a limited number of radio links/channels to the end-users who intend to have a call simultaneously, resulting in capacity limitation. Therefore, the spectrum should be used more efficiently to meet the demanded radio access.

The target of cellular concept is to address these problems. The main idea is simple. Suppose we are planning a radio network for a large city where there are millions of mobile users. Based on the cellular concept, the large area is divided into a number of sub-areas called *cells*. Each cell has its own base station, which is able to provide a radio link for specific numbers of simultaneous users by emitting a low-level transmitted signal. Figure 3.5 illustrates an example of a seven cell cluster of a cellular network.

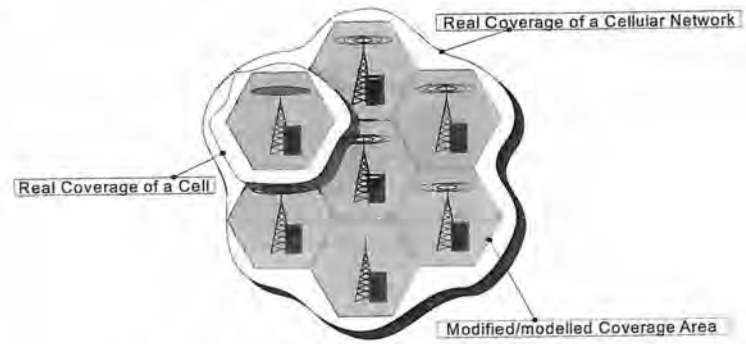


Figure 3.5. A cluster of cells in a cellular network

The nature of radio communication and cellular concept are the primary reasons why the basic architecture of advanced mobile systems is currently what it is. Figure 3.6 illustrates the basic architecture of a cellular concept-based radio system. As shown, the basic architecture of any advanced cellular system consists of base stations, a switching network, and fixed network functionality for backbone transmission.

The cellular concept solution resolves the basic problems of radio systems in terms of radio system capacity constraints, but at the same time it encounters other problems, such as:

- Interference due to the cellular structure, including both inter- and intra-cell interference
- Problems due to mobility
- Cell-based radio resource scarcity

Assume a cellular system with asynchronous users sharing the same radio bandwidth and using the same radio base station in each coverage area or cell, each base station not only receives interference from mobiles in the home cell but also from terminals and base stations located in neighbouring cells. Therefore, depending upon on the source of interference, they can be classified as intra-cell/co-channel, inter-cell/adjacent cell, and interference due to thermal noise.

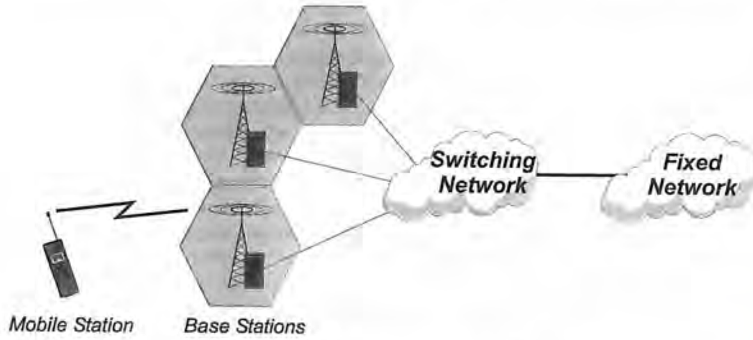


Figure 3.6. Basic structure of a cellular system

Therefore, and as illustrated in Figure 3.7, in order to cope with the total interference received from terminals, interference within the home cell and neighbouring cells should be investigated.

From the mobile station standpoint, inter-cell interference is caused from the signal of base stations or mobile stations of the neighbouring cells of the home cell where the mobile station is camping.

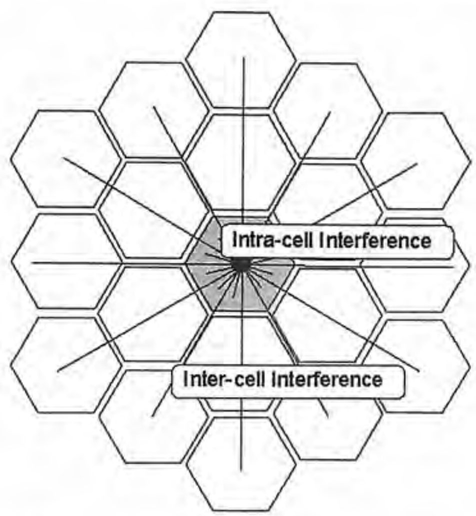


Figure 3.7. [illegible]

In a cellular system, if the same frequency is used in every cell then the mobile station, which is connected to the radio network, encounters the problem of channel interference, including adjacent channel interference and co-channel interference. In addition to that, the network environment may include mountains, hills, and other obstacles, causing multipath problems and path losses. This is solved mainly by utilising the *frequency reuse technique*, i.e. in each cell of cluster pattern a different frequency is used. So, the frequency reuse factor is an essential parameter for radio network planning in any cellular system, which operates based on frequency sharing principles. By optimising the frequency reuse factor the problems of adjacent channel and co-channel hazards can be removed considerably, resulting in increased capacity, which is crucial to the public radio systems. Figure 3.8 shows an example of one-seven and one-one frequency reuse patterns, respectively.

Figure 3.8. Frequency re-use pattern, FR = 7 and FR = 1

The cellular concept increases the radio system capacity especially when utilising with frequency reuse. The smaller the cells, the more efficiently the radio spectrum is used but the cost of the system increases at the same time because more base stations are needed.

Multilayer network designing, including macro-, micro- and pico-cells is a further step in providing advanced system designing solutions in conjunction with the frequency reuse concept to improve the system capacity. It is the subject of network planning to optimise the combination of network structure and frequency reuse to increase the system capacity cost-effectively, avoiding undesirable increase in the number of needed base stations. These solutions, however, set new requirements for the system to handle the mobile's mobility and desirable system performance in terms of quality of services and signalling load.

Although, mobility allows the possibility of being reachable anywhere and any time for the end-user, nevertheless, it sets very strict requirements for the cellular system to support such a feature. Managing the mobile terminals' mobility is one of the most essential parts of any cellular system functionality. Generally, in a radio communication system, paging, location updating, and handover operations provide the user mobility. Handover mechanism guarantees that whenever the mobile is moving from one base station area/cell to another, the radio signal is handed over to the target base station. In addition, location update and paging mechanisms guarantee that the mobile station can be reached even though there is no continuous active radio link between the mobile and corresponding base station. The network always initiates the paging mechanism; mobile station initialises the location update procedure.

3.3. Multiple Access Techniques

The cellular concept approaches the capacity limitation in terms of system coverage. Therefore, it does not alone help the per-cell capacity limitation as far as the simultaneous users are in question. From the radio spectrum standpoint, it is extremely important how the radio resources are allocated to the simultaneous users. Controlling the radio resources has become one of the most critical features of any public radio communication system, which provides services to the huge numbers of service requesters.

Multiple accesses techniques have been developed to combat the problem of simultaneous radio access allocation to the access requesters. The main aspect of any multiple access scheme is how the available frequency band is allocated. The primary problems encountered by multiple access solutions are related to the intrinsic characteristics of radio systems; bandwidth limitation, multipath fading of the radio link and interference from other users in the cellular environment where the cellular concept and frequency reusing have been utilised. The key issue for effective use of the frequency is multiple access. That is, there should be as many users as possible utilising the fixed resource, in this case, frequency.

The multiple access method used in analogue cellular networks is called FDMA (Frequency Division Multiple Access) where every user uses their own frequency (or pair of frequencies) (Figure 3.9).

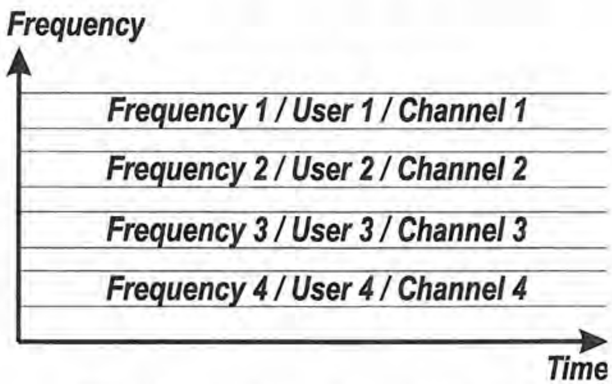


Figure 3.9. FDMA (frequency division multiple access)

In the case of FDMA, having its own frequency or bandwidth specifies the radio channel. When considering the number of radio connections this means one frequency = one user = one channel. This multiple access technique, however, is not an efficient way of sharing the limited radio resources in a public mobile system. Therefore, a more effective way to utilise frequency resources which enables an increase in the system capacity, called TDMA (Time Division Multiple Access) was developed at the earlier phase of radio communication systems. TDMA is the most common multiple access method used in the second generation cellular systems such as in GSM. Also, TDMA based transmission systems are widely used. Maybe the most

In the GSM system Um interface the users of a given frequency divide it in time: every user has a little slice of time (timeslot) for different operations. This timeslot is repeated frequently and this generates an impression of continuous connection. This system makes it possible to have several users – as many as the number of timeslots, which is eight for GSM – simultaneously using the same frequency (Figure 3.10).

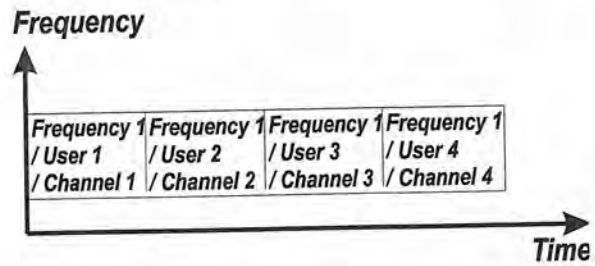


Figure 3.10. TDMA (time division multiple access)

Code Division Multiple Access (CDMA) is another multiple access technique which has been used for similar purposes as FDMA and TDMA. However, it approaches the same problem by utilising a totally different strategy. As a spread-spectrum-based radio access scheme, CDMA is one of the most sophisticated ones, which has been used in different applications. We present a short review of the basic CDMA scheme without going into the subject in detail. Figure 3.11 illustrates the basic radio resource strategy of the CDMA scheme. Unlike in TDMA and FDMA schemes, the radio resource is allocated based on codes in the CDMA scheme. Thus all the simultaneous users can occupy the same bandwidth

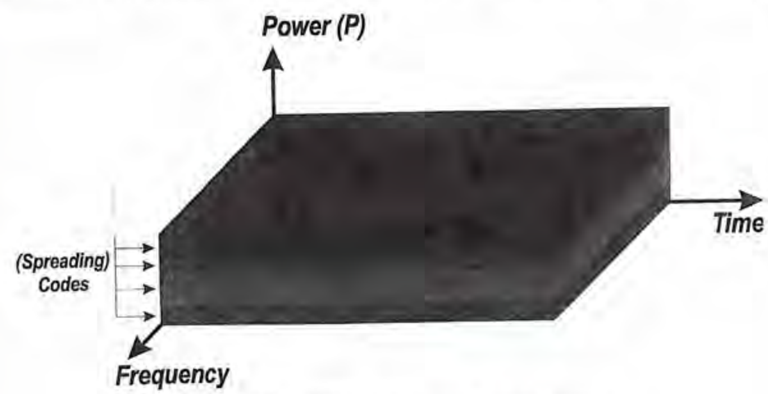


Figure 3.11. CDMA (code division multiple access)

at the same time. Every user is assigned a code/codes varying per transaction and those codes are used for cell, channel and user separation. Every user uses the same frequency band simultaneously and hence there is no "time slots" or frequency allocation in the same sense as in TDMA and FDMA based systems, respectively.

- If the originating bit rate is low, it can be spread well and thus the power required for transmission will be small. This kind of case can be seen as a thin layer in Figure 3.11.
- If the originating bit rate is high it cannot be spread as well and thus the power required for transmission will be higher. This kind of case can be seen as a thick layer in Figure 3.11.

Unlike bandwidth-limited multiple access (FDMA and TDMA) which suffers mainly from co-channel interference due to the high range of frequency reusing, the inter-user uplink interference is the most crucial type of interference in terms of capacity and QoS in the CDMA system. The reason is that inter-user uplink interference increases on a power basis, and the performance of each user becomes poorer as the number of simultaneous users increases in a single cell.

Depending on the spreading signal used in the modulation, the CDMA scheme can be categorised into the following groups:

- Direct Sequence CDMA (DS-CDMA)
- Frequency Hopping CDMA (FH-CDMA)
- Time Hopping CDMA (TH-CDMA)
- Hybrid Modulation CDMA (HM-CDMA)
- MultiCarrier CDMA (MC-CDMA)

In a DS-CDMA scheme, the data signal is scrambled by the user-specific pseudo noise (PN) code at the transmitter side, for example mobile or base station to make the spreading of the signal with the desirable chip rate and process gain. At the receiver (mobile or base station), the original signal is extracted by exactly the same spreading code sequence. As a result, every signal is assumed to spread over the entire bandwidth of the radio connection. Interference may therefore be generated from all directions in contrast to narrow-band cellular systems (Figure 3.12).

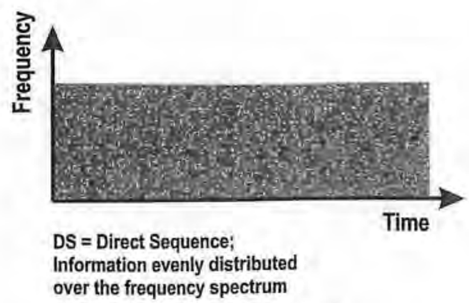


Figure 3.12. DS-CDMA

The frequency reuse for DS-CDMA is one, meaning that all users have the same frequency band to carry their information. The inherent advantage of the DS-CDMA is its multipath fading tolerability. Indeed, in a TDMA-based system like GSM the similar advantage is achieved by utilising a frequency-hopping technique, resembling the spread spectrum scheme. On the other hand, in the case of common shared frequency it is vulnerable to multiuser interference more than TDMA and FDMA.

In FH-CDMA, changing the carrier frequency over the transmitted time of the signal produces the spread bandwidth. That is, during a specific time span, the carrier frequency is kept the same, but after that, the carrier is hopped to another frequency based on the spreading code, resulting in a spread signal (Figure 3.13).

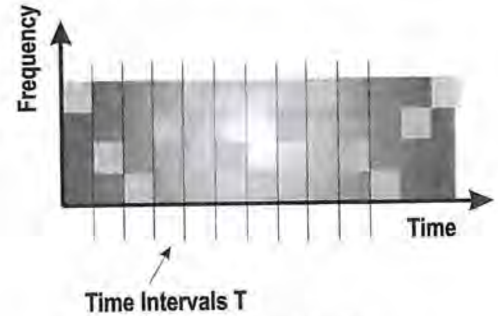


Figure 3.13. FH-CDMA

Depending on the hopping rate of the carrier signal, FH-CDMA is divided into two parts: fast frequency hopping and slow frequency hopping. For the fast frequency hopping modulation scheme, the hopping rate is typically greater than the symbol rate and for the slow frequency scheme, the hopping rate is typically smaller than the symbol rate. Some essential procedures of the CDMA scheme, like power control, are much easier to implement and handle in FH-CDMA than in DS-CDMA. This is partly due to frequency division aspect of this technique when applying to the cellular systems. On the other hand, producing a high-rate frequency hopping is, however, a complex issue in FH-CDMA.

In the time hopping CDMA scheme, the transmitted signal is divided into frames, and those frames are further divided into time intervals. During the transmission time, the data burst is hopped between frames basically by utilising code sequences. Principally, the aim is to choose the code sequence so that the simultaneous signal transmission in the same frequency may be minimised. The most important idea behind this approach is to create synergy between the strengths of the previously described schemes to provide convenient schemes for specific applications. The combinations of these CDMA schemes leads to different hybrid schemes such as DS/FH, DS/TH, FH/TH, and DS/FH/TH.

Unlike DS-CDMA, in MC-CDMA the entire frequency band is used with several carriers instead of one. Therefore, the MC-CDMA transmitter basically spreads the original data over different sub-carriers using different frequency bands by denoting a spreading code in the frequency demand (Figure 3.14). The MC-CDMA schemes are mainly categorised into two groups. One spreads the original data stream using a given spreading code, and then modulates different sub-carriers. The other one is basically similar to DS-CDMA.

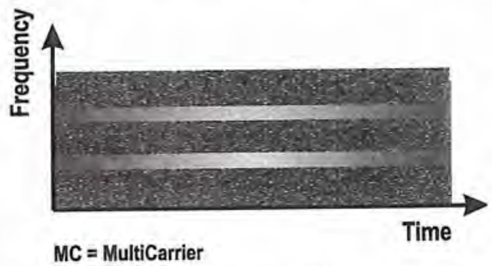


Figure 3.14. MC-CDMA

3.4. Regulation

Conventionally, regulation has been an inseparable part of the development in communication systems. The role of regulation is concretised in two main areas, radio spectrum allocation and overall technical standardisation of the communication systems. In both areas, the main drivers of the regulation rely on an technology–engineering basis and economic–political aspects. Legal administration plays an important role in supporting the entire regulation process (Figure 3.15).

Figure 3.15. Interaction of different factors in the regulation process

So far, we have described the different techniques applied in radio systems to utilise the radio spectrum as efficiently as possible. Unlike the public understanding, the radio spectrum is not an unlimited natural resource but rather it is extremely scarce. Although, the advanced systems exploit the available radio spectrum more efficiently, nevertheless, rapid growth in ... the scarcity of the spectrum. Therefore, it is essential

scarce natural resource. Indeed, radio spectrum regulation has always been a cornerstone of the communication business. In order to understand the role of radio spectrum regulation two main aspects should be distinguished, technical–engineering and economic–political aspects. Regarding the technical aspect of the radio spectrum, spectral efficiency is of primary concern in designing of any wireless system. As we mentioned earlier in this chapter, many technical approaches, i.e. multiple accesses, cellular concept, etc. have been developed to use the radio spectrum as efficient as possible.

Figure 3.16 illustrates the electromagnetic spectrum, including the radio wave spectrum from 3 kHz up to 300 GHz. On the other hand, the figure shows that most of the allocated radio spectrum allocated for wireless applications ranges from 100 to over 2000 MHz. The reason is partly because of the characteristics of the radio wave, that is, the higher the radio frequency the higher the radio wave attenuation. Therefore, technically, the higher the radio frequency the system implementation is more complex and non-cost-effective. In a cellular system, higher frequency application may increase the number of base stations needed in the subsystem for providing the required coverage.

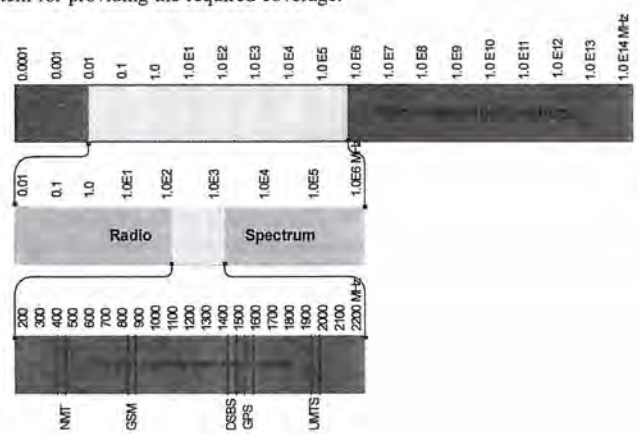


Figure 3.16. The electromagnetic spectrum, radio wave portion and the most demanded portion

The economic–political aspects of the radio spectrum are still more complicated. The general national radio administrations are in charge of careful regulation of the whole radio spectrum. In addition, many international organisations are involved in radio spectrum regulation. The basic goal is to make the scarce radio resource available, in a manner that, on one hand, the system suppliers and users are content, and on the other hand, the spectrum has been used efficiently.

In addition to these factors, the communication market environment is extremely competition-oriented and continuously changing. Thus, spectrum regulation has also become a key competitive factor both nationally and globally and hence the process of deregulation is aligned, resulting in divergent spectrum allocation. For instance, a radio spectrum allocated to military systems has been a high priority in conjunction with international joint spectrum allocation strategy.

Due to dramatic growth in the wireless communication market, it seems that the scarcity of the radio spectrum will be more crucial in the future. Therefore, it is essential to make a trade-off between technical solutions, wireless applications, and the regulation to obtain a fair level of compromise solution. In terms of 1G, the radio spectrum was something to be given away free-of-charge except in the US, where radio spectrum slots have always been subject to regulation fees. The free-of-charge radio spectrum principle was going to be continued in 2G and this development was partially successful.

The spectrum regulation for the 3G system is bringing more competition to the field of communication technology and business. It is also bringing a common understanding that the radio frequency is not an endless natural resource as the number of mobile users is dramatically increasing and this technology is converging with the Internet. The radio spectrum regulation fees increased to a very high level. 3G spectrum regulation aspects are a complete story of their own and estimating their consequences is not within the scope of this book.

3.5. Essentials of the 3G Radio Path

3.5.1. Frequency Band and Regulatory Issues

3G spectrum regulation has been a hot subject of discussion during the last few years worldwide. It includes many sophisticated issues, which vary greatly between countries and operators. Firstly, the needed spectrum room for 3G is not available similarly in different countries. That is basically because of the spectrum portion occupied by other systems already in use or the guard bands needed for the systems, which operate on adjacent frequency bands. Secondly, the political approaches for spectrum licensing also vary from country to country. Naturally, operators involved in 2G systems are willing to have the edge technology by participating in 3G development work. In addition, many Greenfield operators (those newcomers to the mobile communications business) intend to reinforce their rule in the market by launching the 3G system. Therefore, the 3G spectrum regulation environment demands close international and national co-operation in order to harmonise spectrum allocation world-wide and manage it efficiently within countries (Figure 3.17).

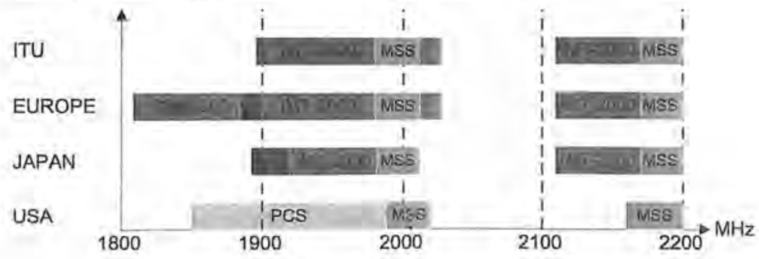


Figure 3.17 The 3G system spectrum

From a 3G point of view, it has been decided (in mid-1999 by OHG) that there will be three CDMA variants in use. These are:

- DS-WCDMA-FDD: Direct Sequence – Wideband Code Division Multiple Access – Frequency Division Duplex
- DS-WCDMA-TDD: Direct Sequence – Wideband Code Division Multiple Access – Time Division Duplex
- MC-CDMA: Multi Carrier – Code Division Multiple Access

In these terms, the first part describes the information spreading method over the frequency spectrum, the second part indicates multiple access scheme and the last part expresses how the different transmission directions (uplink and downlink) are separated. Before proceeding it should be noted that the word "Wideband" here does not have any specific meaning. Originally it was inserted to the name because the Euro-Japanese version of CDMA used wider bandwidth than the American one. The wider band makes it possible to insert some attractive features to the system like, for instance, multimedia services with adequate bandwidth and macro diversity. Due to the OHG decision, however, all the three CDMA variants use similar bandwidth.

The WCDMA-FDD uses frequencies of 2110–2170 MHz downlink (from the BS to the UE) and 1920–1980 MHz uplink (from the UE to the BS). Air interface transmission directions are separated by different frequencies and duplex distance is 190 MHz.

The TDD variant of the WCDMA uses a frequency band located in both sides of the WCDMA-FDD uplink. The lower frequency band offered for the TDD variant is 20 MHz and the higher one is 15 MHz.

For the purpose of comparison it should be mentioned that the GSM1800 system uses frequencies of 1805–1880 MHz downlink (from the BTS to the MS) and 1710–1785 MHz uplink (from the MS to the BTS). Air interface transmission directions are separated from each other by different frequencies and duplex distance is 95 MHz.

In this book we will shortly introduce the DS-WCDMA-FDD basics and related technology, since it is the first radio access technology to be used when implementing UMTS wideband radio access. Due to simplicity, we further use the name "WCDMA" as a synonym of "DS-WCDMA-FDD".

3.5.2. Basic Concepts

The principles of the WCDMA technique are based on spread spectrum. Therefore, in order to outline the WCDMA scheme, it is essential to understand the general principles of spread spectrum. We next describe briefly the highlights of the spread spectrum scheme.

Spread spectrum is a well-proven modulation scheme that creates a bandwidth for the transmitted signal much wider than the bandwidth of the actual information, which is intended for transmission. The history of the spread spectrum technique goes back almost to the 1950s since the use of the scheme for military communication become a reality. Thanks to the efficient use of the radio spectrum by allowing additional users to share the same frequency and utilising efficient mechanisms like fast power control, diversity and soft handover the CDMA variant of the spread spectrum scheme has become a very promising radio access alternative for public applications of mobile communication systems. The main advantages of the spread spectrum scheme can be summarised as follows:

- Its resistance to radio interference and jamming
- It lowers the probability of interception by an adversary
- It is resistant to signal interference from multiple transmission signal branches
- It provides multiple access facility with a reuse factor equal to one
- It supports the means for measuring range, or the distance between two points
- It yields the possibility of utilising diversity techniques, including multipath diversity, as well as frequency and time diversity.
- It provides user access at any time without waiting for a free channel as far as the level of interference meets the system's tolerance.

Having outlined the main characteristics of the spread spectrum scheme, we further present the basic modulation technique of the scheme. Figure 3.18 illustrates the main principles of the modulation process of the DS spread spectrum scheme. Assume that the radio signal is transmitted from the base station to the mobile station. At the base station, the transmitted signal with rate R is spread by convoluting with a wideband spreading signal, creating a spread signal with bandwidth W. At the mobile, the received signal is multiplied by exactly the same spreading signal. Now if the spreading signal, locally generated at the mobile, is synchronised with the spread code/signal, the result is the original signal plus possibly some trumped up higher frequency components which are not part of the original signal, and hence can easily be filtered. If there is any undesired signal at the mobile, on the other hand, the spreading signal will affect it just as it did the original signal at the base station, spreading it to the bandwidth of the spreading signal.

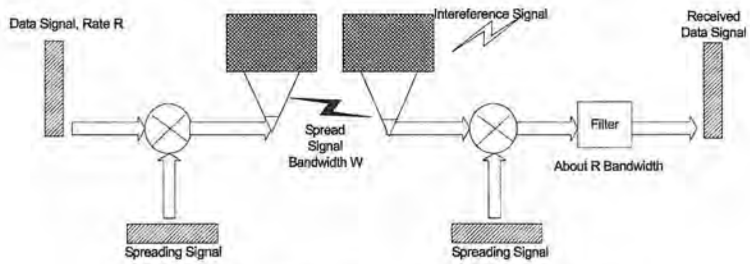


Figure 3.18. The basic technique of DS spread spectrum

In WCDMA, the data stream of the base station transmitter handles data in the downlink direction which represents the traffic from the network to the terminal. This traffic uses several channels in the Uu interface. In the Uu interface the effective bandwidth for WCDMA is 3.84 MHz and with guard bands the required bandwidth is 5 MHz (Figure 3.19).

As mentioned earlier, in a DS-CDMA scheme, the data signal is scrambled by utilising the user-specific PN code at the transmitter side to achieve the spreading signal and at the receiver the received signal is extracted by using the same code sequence. In order to get a better understanding of the issue, we briefly discuss the *Information Theory*.

Information Theory is a mathematical model explaining the principles of signal transfer. This theory does not make any difference between different technologies involved because

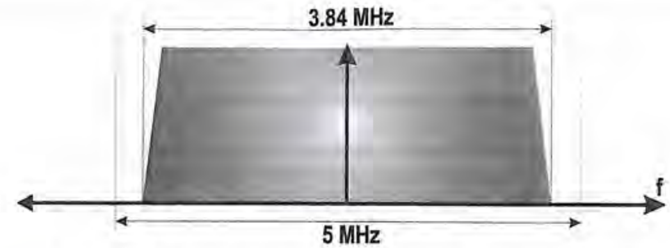


Figure 3.19. One direction of the WCDMA carrier and it's dimensions

the basic mathematics is always the same. In this context, there are several simplified principles, which can be extracted out from the Information Theory:

- The information to be transferred presents certain power, say, P_{inf}
- The wider the band for the information transfer, the smaller is the power presenting transferred information in a dedicated (small) point within the information transfer band. In other words, the total power P_{inf} is an integral over the information transfer band in this case.
- The more information there is to be transferred, the more power is required. Thus, when the power increases momentarily, P_{inf} increases, too. In this context, the higher the original bit rate to be transferred, the more power required.

If we take this into account and combine the information presented in Figure 3.11, we are able to illustrate how WCDMA treats a single originating piece of user data called a bit.

In the WCDMA air interface, every originating information bit is like a "box" having constant volume but the dimensions of the "box" change depending on the case. Referring to Figure 3.20, the depth of the "box" (frequency band) is constant in the WCDMA. The other

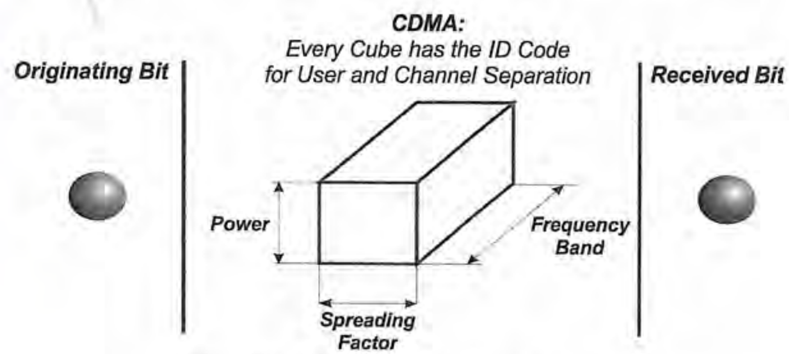


Figure 3.20. WCDMA air interface and bit treatment

two dimensions, power and spreading factor are subject to change. Based on this the following conclusions can be made:

- The better the signal can be spread, the smaller is the required energy per bit (power). This can be applied if the originating bit rate is low. In other words, spreading factor increases and power decreases.
- The smaller the spreading factor, the more energy is required per bit (power). This is applied when the originating bit rate is high. In other words, spreading factor increases and power decreases.

A confusing issue with WCDMA is that a "bit" is not a "bit" in all cases. The term bit refers to the information bit, which is a bit occurring in the original user data flow. The bit occurring in the code used for spreading is called chip. Based on this definition, we are able to present some basic items needed in WCDMA.

The bit rate of the code used for original signal spreading is, as defined, 3.84 Mb/s. This value is constant for all WCDMA variants used in 3G networks. This is called *System Chip Rate* and it is expressed as 3.84 Mcps (mega chips per second). With this system chip rate the size of one chip in time is 1/3 840 000 = 0.00000026041s.

As mentioned previously, the basic idea in WCDMA is that the signal to be transferred over the radio path is formed by multiplying the original, baseband digital signal with another signal, which has a much greater bit rate. Because both of the signals consist of bits, one must make a clear separation of the kind of bits in question (Figure 3.21). Hence:

- In the air interface the information is transmitted as *symbols*. Symbol flow is a result of modulation. Before modulation the user data flow consisting of bits has gone through channel coding, convolutional coding and rate matching. Referring to Figure 3.20, the cube in the middle of the picture actually represents 1 symbol. Depending on the used modulation method 1 symbol represents different amount of bits. In the case of DS-WCDMA-FDD, 1 symbol transmitted in an uplink direction represents 1 bit and 1 symbol

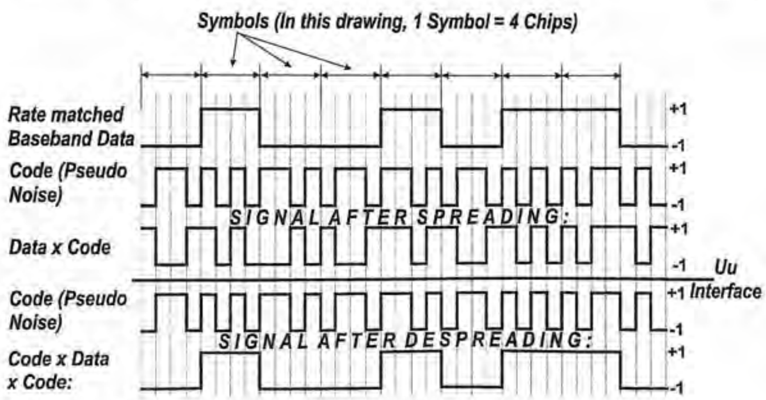


Figure 3.21 Bit, chip and symbol

transmitted in downlink direction represents 2 bits. This difference is due to the different modulation methods used in uplink and downlink directions.

- One bit of the code signal used for signal multiplying is called a *chip*.

But how can the desired signal be captured at the receiver side? In principle, the process is quite straightforward; every receiver uses its unique code to pick up the desired signal. The received signal is multiplied by a receiver-specific code, resulting in the multiplication data. If the right code and desired received signal are multiplied, the result will be the post-integrated data with the clear peaks on the signal, otherwise, the post-integrated data does not include clear signal peaks to be processed further.

The spreading factor is a multiplier describing the number many chips is used in the WCDMA radio path per 1 symbol. The spreading factor K can be expressed mathematically as follows:

$$K = 2^k$$

where $k = 0, 1, 2, \ldots, 8$.

For instance, if $k = 6$ the spreading factor K gets a value of 64 indicating that 1 symbol uses 64 chips in the WCDMA radio path in the uplink direction (refer to Tables 3.1 and 3.2). Another name for spreading factor is *processing gain* (G_p), and it can be expressed as a function of used bandwidths:

$$G_p = \frac{B_{Uu}}{B_{Bearer}} = \frac{\text{System chip rate}}{\text{Bearer bit rate}} = \text{Spreading factor}$$

In the formula, B_{Uu} stands for the bandwidth of the Uu Interface and B_{Bearer} is the bandwidth of the rate-matched baseband data. In other words, B_{Bearer} contains already excessive information like channel coding and error protection information. Based on the equation above and taking into account the different bit amounts 1 symbol carries in uplink and downlink directions we are able to calculate the bearer bit rates available in WCDMA.

These figures are indicative since the share of user data (payload) changes according to the used radio channel configuration.

The WCDMA system uses several codes. In theory, one type of code should be enough but in practise, the radio path physical characteristics require that the WCDMA system should use

Table 3.1 Spreading factor–symbol rate–bit rate relationship in the uplink direction

Spreading factor value	Symbol rate (ks/s)	Channel bit rate (kb/s)
256	15	15
128	30	30
64	60	60
32	120	120
16	240	240
8	480	480
4	960	960

Table 3.2 Spreading factor–symbol rate-bit rate relationship in the downlink direction

Spreading factor value	Symbol rate (ks/s)	Channel bit rate (kb/s)
512	7.5	15
256	15	30
128	30	60
64	60	120
32	120	240
16	240	480
8	480	960
4	960	1920

different codes for different purposes, and those codes should have certain features making them suitable for their use. There are basically three kinds of codes available, channelisation codes, scrambling codes and spreading code(s). Their use is shown in Table 3.3.

Table 3.3 Code types of WCDMA

	Uplink direction	Downlink direction
Scrambling codes	User separation	Cell separation
Channelisation codes	Data and control channels from the same terminal	Users within one cell
Spreading code	Channelisation code × scrambling code	Channelisation code × scrambling code

Another confusing issue with WCDMA is that the same item could have many names. This naming confusion is also present as far as codes are concerned. Scrambling code, for instance, is also known as gold code (often used in radio path related technical publications) and long code. From these alternatives, the name scrambling code is the preferred name. The scrambling code is used in the downlink direction for cell/sector separation. Scrambling codes are used also in the uplink direction. In this case, the users (mobiles) are separated from each other with this code.

In the case of downlink, a total of $2^{18} - 1 = 262\ 143$ scrambling codes, numbered 0…262 142 can be generated. However not all the scrambling codes are used. The scrambling codes are divided into 512 sets each having a primary scrambling code and 15 secondary scrambling codes (Figure 3.22).

The primary scrambling codes consist of scrambling codes $n = 16*i$ where $i = 0,\ldots,511$. The ith set of secondary scrambling codes consists of scrambling codes $16*i + k$, where $k = 1,\ldots,15$. There is a one-to-one mapping between each primary scrambling code and 15 secondary scrambling codes in a such a way that the ith primary scrambling code corresponds to ith set of scrambling codes.

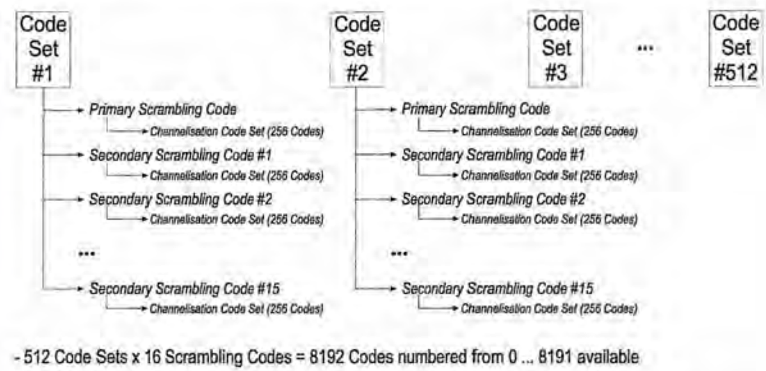


Figure 3.22. Scrambling and channelisation codes; amounts and relationship

Hence, according to the above, there are $k = 0, 1,\ldots,8191$ scrambling codes available. Based on 3GPP TS 25.213,each of these codes is associated with an even alternative scrambling code and an odd alternative scrambling code that may be used for compressed frames. The even alternative scrambling code corresponding to scrambling code k is scrambling code number $k + 8192$, while the odd alternative scrambling code corresponding to scrambling code k is scrambling code number $k + 16384$. The set of primary scrambling codes is further divided into 64 scrambling code groups, each consisting of eight primary scrambling codes. The jth scrambling code group consists of primary scrambling codes $16*8*j + 16*k$, where $j = 0,\ldots,63$ and $k = 0,\ldots,7$.

Each cell is allocated one and only one primary scrambling code. The primary CCPCH (Common Control Physical Channel) is always transmitted using the primary scrambling code. Hence, the other downlink physical channels can be transmitted with either the primary scrambling code or a secondary scrambling code from the set associated with the primary scrambling code of the cell.

For the uplink case the situation is different because there are millions of uplink scrambling codes available. The specified amount of uplink scrambling codes for WCDMA is 2^{24}. All uplink channels shall use either short or long scrambling codes, except the PRACH (Physical Random Access Channel), for which only the long scrambling code is used. Therefore, uplink code allocation is not as crucial as downlink code allocation in WCDMA.

Channelisation codes are used for channel separation both in uplink and downlink directions. Channelisation codes are the ones that have different spreading factor values and thus also different symbol rates. There is a total of 256 pieces of channelisation codes available and the spreading factor indicates how many bits of those codes are used in the connection. Thus, the greater the spreading factor value is, the better the channelisation codes are utilised and the radio resources are used in an optimal way. In the case of high user data rates the spreading factor gets a relatively small value. Respectively, this leads to the situation where high user data rates consume more air interface code capacity.

As is their nature, the channelisation codes are *orthogonal*, or, one can say, they have orthogonal properties. Orthogonality as a term means that the channelisation codes in the 256-member code list are selected in a way that they interfere with each other as little as possible. This is necessary in order to have a good channel separation. On the other hand, the code used for user and cell separation (scrambling code) must have good correlation properties. The channelisation codes do have those and this is the basic reason why both scrambling and channelisation codes are used.

Every WCDMA cell uses normally one downlink scrambling code, which is locally unique and acts basically like a cell ID. The characteristics of this scrambling code are pseudo-random, i.e. it is not always orthogonal. Under this scrambling code the cell has a set of channelisation codes, which is orthogonal in nature and used for channel separation purposes.

In WCDMA, the users transmitting and receiving use the whole available frequency band simultaneously in time. To separate different transmissions spread over the frequency band, spreading codes are used. A spreading code is a unique code assigned to the beginning of the transaction by the network. A spreading code can be imagined to be like a "key" which is used both by the mobile and the network. Both ends of the connection use this "key" to open the noise-like transmitted wideband signal. Or to be exact, to extract the correct wideband transmission away from the frequency band, since the transmitted wideband signal may contain many mobile-network connections.

From the point of view of the spreading code, the capacity of a cell depends on the downlink scrambling code amount assigned for the cell (minimum is 1). Every downlink scrambling code then has a set of channelisation codes under it and every call/transaction requires one channelisation code to operate. In practise, one spreading code is actually *scrambling code × channelisation code*. If channelisation codes are not used, the spreading code is the same as the scrambling code. Also, the spreading code depends on the information type to be delivered. The information being common in nature does not use channelisation code. For example, the broadcast information the cell delivers downlink is one of those.

3.5.3. WCDMA Radio Channels

The WCDMA radio access allocates bandwidth for users and the allocated bandwidth and its controlling functions are handled with the term "Channel". The functionality implemented through the WCDMA defines what kinds of channels are required and how they are organised. The channel organisation the WCDMA uses is a three-layer one; there are logical channels, transport channels and physical channels. From these, the logical channels describe the types of information to be transmitted, transport channels describe how the logical channels are to be transferred and the physical channels are the "transmission media" providing the radio platform through which the information is actually transferred (Figure 3.23).

Referring to the architecture issues explained in Chapter 1, the channel structures and their use differs remarkably from the GSM. The term physical channels means different kinds of bandwidths allocated for different purposes over the Uu interface. In other words, the physical channels actually form the physical existence of the Uu interface between the UE domain and access domain. In GSM the physical channels and their structure is recognised by the BSC but in WCDMA the physical channels really exist in the Uu interface and the RNC is not necessarily aware of their structure at all.

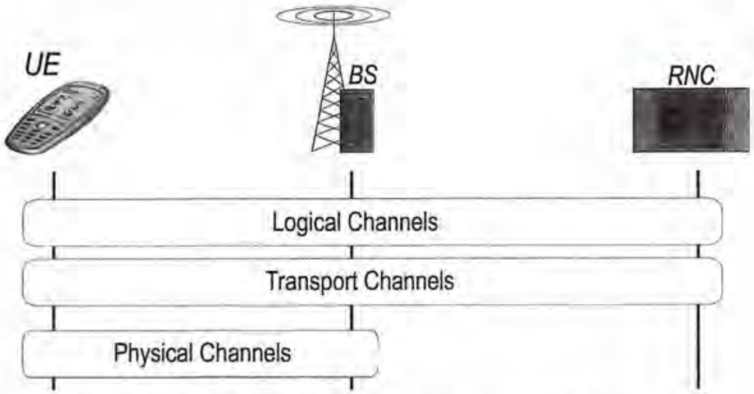


Figure 3.23. Channel types and their location in UTRAN

Instead of physical channels the RNC "sees" transport channels. Transport channels carry different information flows over the Uu interface and the physical element mapping these information flows to the physical channels is the BS (Base Station). Logical channels are not actually channels as such, rather they can be understood as different tasks the network and the terminal should perform in different moments of time. These partially timely structures are then mapped into transport channels performing actual information transfer between the UE domain and access domain.

Concerning the logical channels, the UE and the network have different tasks to do. Thus, the logical, transport and physical channel structures are a bit different in either direction. Roughly, the network has the following tasks to perform:

- The network must inform the UE about the radio environment. This information consists of, for instance, the code value(s) used in the cell and in the neighbouring cells, allowed power levels, etc. This kind of information the network provides for the UE through the logical channel called the Broadcast Control Channel (BCCH).
- When there is a need to reach a UE for communication (for instance, a mobile terminated call) the UE must be paged in order to find out its exact location. This network request is delivered on the logical channel called the Paging Control Channel (PCCH).
- The network may have certain tasks to do, which are or may be common for all the UE residing in the cell. For this purpose the network uses the logical channel called Common Control Channel (CCCH). Since there could be numerous UE using the CCCH simultaneously, the UE must use U-RNTI (UTRAN Radio Network Temporary Identity) for identification purposes. By investigating the received U-RNTI the UTRAN is able to route the received messages to the correct serving RNC. U-RNTI is discussed in Chapter 4.
- When there is a dedicated, active connection, the network sends control information concerning this connection through the logical channel called Dedicated Control Channel (DCCH).

- Dedicated traffic: the dedicated user traffic for one user service in the downlink direction is sent through the logical channel called Dedicated Traffic Channel (DTCH).
- Common Traffic Channel (CTCH) is an unidirectional channel existing only in the downlink direction and it is used when transmitting information either to all UE or a specific group of UE in the cell.

The transport channels shown in Figure 3.24 except one are mandatory ones. The mandatory transport channels are Broadcast Channel (BCH), Paging Channel (PCH), Forward Access Channel (FACH) and Dedicated Channel (DCH). In addition to these, the operator may configure the UTRA to use the Downlink Shared Channel (DSCH). In these transport channels, the only dedicated transport channel is DCH; the others are common ones. In this context the term "dedicated" means that the UTRAN has allocated the channel to be used between itself and certain terminals. The term "common" means that several terminals could use the channel simultaneously.

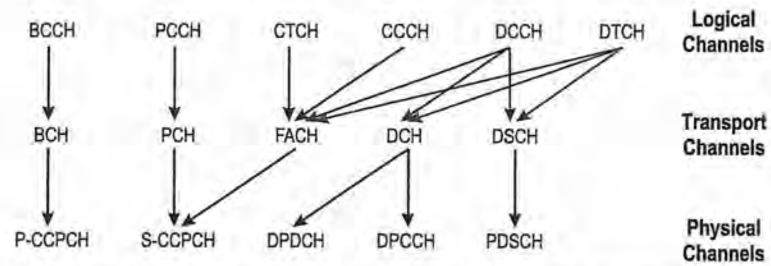


Figure 3.24. Logical, transport and physical channel mapping in the downlink direction

The BCH carries the content of BCCH, i.e. the UTRA specific information to be delivered in the cell. This information consists of, for example, random access codes, access slot information and information about the neighbouring cells. The UE must be able to decode the BCH in order to register to the network. The BCH is transmitted with relatively high power because every terminal in the intended cell coverage area must be able to "hear" it. The PCH carries paging information. The PCH is used when the network wishes to initiate a connection with certain UE. FACH carries control information to the UE known to be in the cell. For example, when the RNC receives a random access message from the terminal, the response is delivered through FACH. In addition to this, the FACH may carry packet traffic in the downlink direction. One cell may contain numerous FACH but one of these is always configured in such a way (with low bit rate) that all the terminals residing in the cell area are able to receive it. The DCH carries dedicated traffic and control information, i.e. the DCCH and DTCH. It should be noted that one DCH might carry several DTCH depending on the case. For example, a user may have a simultaneous voice call and video call active. The voice call uses one logical DTCH and the video call requires another logical DTCH. Both of these, however, use the same DCH. From the UTRA capacity point of view, the aim is to use common transport channels as much as possible, since the dedicated channels will occupy radio resources. The optional

DSCH is a target of increasing interest. It carries dedicated user information (DTCH and DCCH) for packet traffic and several users can share it. In this respect it is better than DCH, since it saves packet traffic related network resources in the downlink direction. Another point is that the maximum bit rate for DSCH can be changed faster than in DCH. The expected wide use of services producing occasional packet bursts, like for instance web surfing, have increased the interest towards DSCH.

In the uplink direction the logical channel amount required is smaller. There are only three logical channels, CCCH, DTCH and DCCH. These abbreviations have the same meaning as in the downlink direction (Figure 3.25).

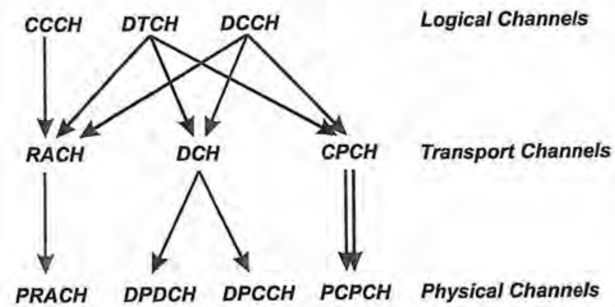


Figure 3.25. Logical, transport and physical channel mapping in the uplink direction

There are three mandatory transport channels in the uplink direction being Random Access Channel (RACH), DCH and Common Packet Channel (CPCH). The RACH carries control information from the UE to the UTRAN, like for instance, connection set-up requests. In addition to this the RACH may carry small amounts of packet data. The DCH is the same as in the downlink direction, i.e. dedicated transport channel carrying DCCH and DTCH information. The CPCH is a common transport channel meant for packet data transmission. It is a kind of extension for RACH and the CPCH counterpart in the downlink direction is FACH.

When the information is collected from the logical channels and organised to the transport channels it is in ready-to-transfer format. Before transmitting the transport channels are arranged to the physical channels. The Uu interface contains more physical channels than Figs. 24 and 25 indicate. The other physical channels present are used for physical radio media control, modification and access purposes. The overview of WCDMA physical channels is provided in Figure 3.26.

The physical channels are used between the terminal and the base station. Due to the network architecture solution explained at the beginning of this book, the physical access (i.e. the physical channels) is separated from the other layers. This arrangement makes it possible to swap the physical radio access medium below the other layers, in theory. In practise the radio access medium change will reflect on higher layers but this arrangement minimises those changes.

The Primary Common Control Physical Channel (P-CCPCH) carries the BCH in the downlink direction. The P-CCPCH is available in a way that all the terminals populated

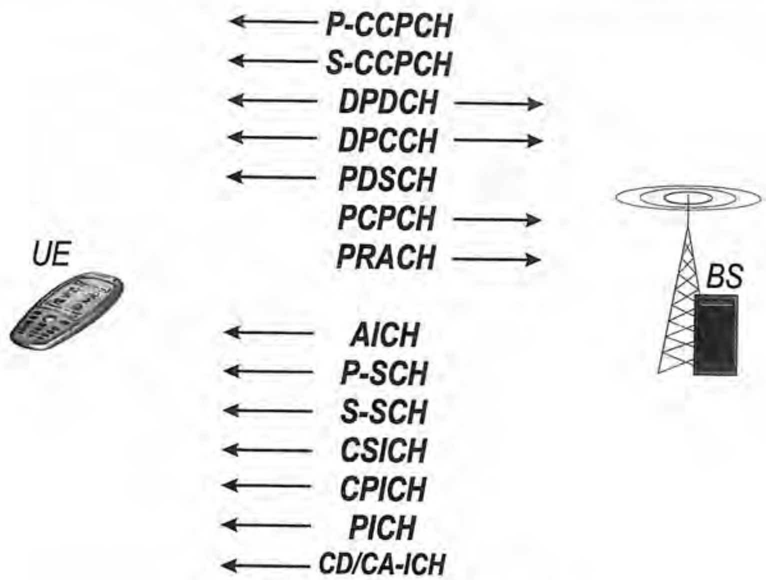


Figure 3.26. WCDMA physical channels

within the cell coverage are able to demodulate its contents. Because of this requirement, the P-CCPCH has some characteristics, which actually are limitations when compared to the other physical channels in the system. The P-CCPCH uses a fixed channelisation code and thus its spreading code is fixed, too. This is a must because otherwise the terminals are not able to "see" and demodulate the P-CCPCH. The P-CCPCH bit rate is 30 kb/s with a spreading factor value of 256. The bit rate must be low because this channel is transmitted with relatively high power. If using higher bit rates, the interference starts to increase thus limiting the system capacity.

The Secondary Common Control Physical Channel (S-CCPCH) carries two transport channels in it: Paging Channel (PCH) and Forward Access Channel (FACH). These transport channels may use the same or separate S-CCPCH, thus a cell always contains at least one S-CCPCH. The bit rate of S-CCPCH is fixed and relatively low due to the same reasons concerning the P-CCPCH. At a later phase the S-CCPCH bit rate can be increased by changing system definitions. The configuration of S-CCPCH is variable: depending on the case, the S-CCPCH can be configured differently in order to optimise the system performance. For instance, the pilot symbols can be included or not. Referring to the variable configuration alternatives of the S-CCPCH, one alternative to increase system performance is to multiplex the PCH information together with the FACH to the S-CCPCH and the PCH related paging indications to a separate physical channel called a Paging Indicator Channel (PICH).

The Dedicated Physical Data Channel (DPDCH) carries dedicated user traffic. The size of the DPDCH is variable and it may carry several calls/connections in it. As the name says, it is a dedicated channel, which means that it is used between the network and *one* user. The dedicated physical channels are always allocated as pairs for one connection: one channel for control information transfer and the other for actual traffic. The DPCCH transfers the control information during the dedicated connection. Figure 3.27 shows how the DPDCH and the DPCCH are handled in both uplink and downlink directions.

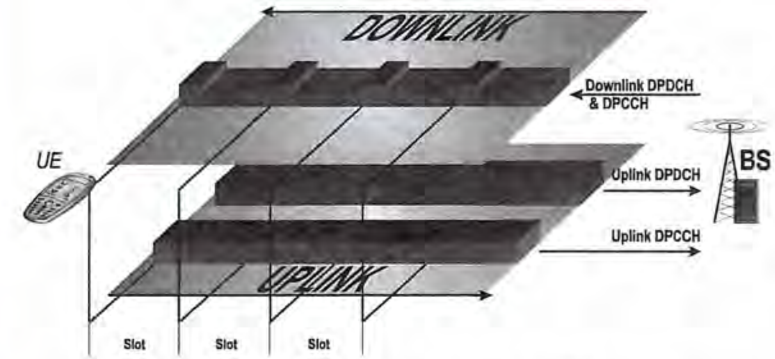


Figure 3.27. DPDCH and DPCCH in uplink and downlink directions

In the downlink direction the DPDCH (carrying user data) and DPCCH (carrying for example power control data rate information) are time multiplexed. If there is nothing to carry in DPDCH the transmitted signal is pulse-like causing EMC type disturbances, this is not a problem in the downlink direction. In the uplink direction the DPDCH and DPCCH are separated by I/Q modulation. If there is no user data to carry in DPDCH, no pulse-like disturbances exist either. The outcome of the I/Q modulation in the UE is actually one channel but carrying two information branches (see Figure 3.27) and both of the information branches consume one code resource.

Together, the DPCCH and the DPDCH carry the contents of the transport channel DCH. When the dedicated connection uses a high peak bit rate the system easily starts to suffer from the lack of channelisation codes in the cell. In this case, there are two options: either to add a new scrambling code to the cell or to use common channels for the dedicated data transmission. The adding of the scrambling codes is not recommended because the orthogonality is lost. Instead, using common channel resources for packet data transmission is seen a better way to increase capacity. The downlink DCH is able to provide information whether the receiving UE must decode the Physical Downlink Shared Channel (PDSCH) for additional user information. The PDSCH carries the DSCH transport channel and as it was explained earlier, the DSCH is an optional feature the operator may configure to be used or not.

If there is a need to send packet data in an uplink direction and the RACH packet transfer capacity is not enough, the UE may use CPCH (uplink Common Packet Channel). The correspondent physical channel in the uplink direction is PCPCH (Physical (uplink) Common Packet Channel). The counterpart of the CPCH in the downlink direction is DPCCH.

The PRACH carries information related to the Random Access Procedure (Figure 3.28). With this procedure the terminal accesses to the network and also small data amounts can be transferred. The Random Access Procedure has the following phases:

1. The UE decodes the BCH information on the P-CCPCH and finds out the RACH slots available as well as the scrambling code(s).
2. The UE selects randomly one RACH slot to be used.
3. The UE sets the initial power level to be used (this is based on the received downlink power level) and sends so-called preamble to the network.
4. The terminal decodes the AICH (Acquisition Indication Channel) to find out if the network notified the sent preamble. If it was not the UE sends preamble again but with a higher bit power level.
5. When the AICH indicates that the network notified the preamble, the UE sends RACH information on the PRACH. The length of the RACH information sent is either one or two WCDMA frames, in time this takes either 10 or 20 ms.

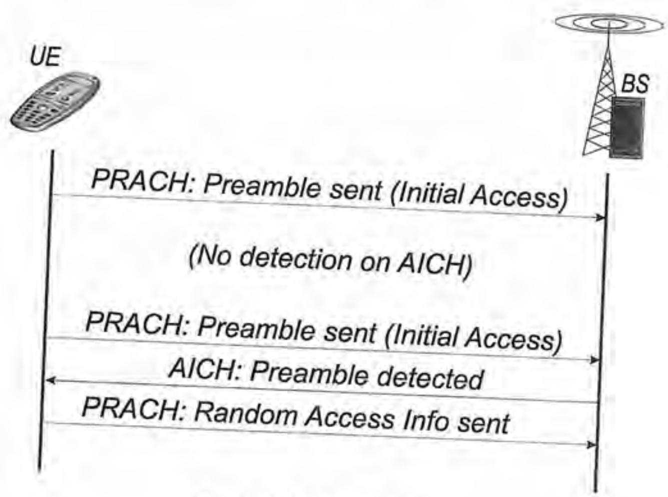


Figure 3.28. Random access

The synchronisation channel provides the cell search information for the UE residing the cell coverage area. The SCH is actually a combination of two channels, Primary Synchronisation Channel (P-SCH) and Secondary Synchronisation Channel (S-SCH). The P-SCH uses a fixed channelisation code, whose length is 256 and this channelisation code is the same in every cell of the system. When the UE has demodulated the P-SCH it has achieved frame and

slot synchronisation of the system and it also knows to which scrambling code group the cell to be accessed belongs.

The Common Pilot Channel (CPICH) is an unmodulated code channel which is scrambled with the cell specific scrambling code. The CPICH is used for dedicated channel estimation (by the terminal) and to provide channel estimation reference(s) when common channels are concerned. In this respect, the pilot signal is somewhat the same with the same functionalities as the training sequence included in the middle of a GSM burst. Normally, one cell has only one CPICH but there can be two of them. In this case, those channels are called primary CPICH and secondary CPICH. The cell may contain the secondary CPICH, for instance, when the cell contains narrow beam antenna aiming to offer service for a dedicated "hot spot" area. Thus, a dedicated area uses the secondary CPICH and the primary CPICH offers pilot for the whole cell coverage area. The terminals listen to the pilot signal continuously and this is why it is used for some "vital" purposes in the system, e.g. handover measurements and cell load balancing: From the system point of view, the CPICH power level adjustment balances load between cells: the UE always searches the most attractive cells and by decreasing the CPICH power level the cell is less attractive.

The other physical channels listed in Figure 3.26 are CPCH Status Indication Channel (CSICH), CPCH Collision Detection Indicator Channel (CD-ICH) and CPCH Channel Assignment Indicator Channel (CA-ICH). The CSICH uses free space occurring in AICH and the CSICH is used to inform the UE about CPCH existence and configuration. To avoid collisions (two UE using same identity patterns), a CD-ICH and CA-ICH are used. These physical channels transfer the collision detection information towards the UE.

3.5.4. WCDMA Frame Structure

In order for the radio access to handle the control actions like timing, synchronisation arrangements, transmission assurance, etc. between the network and mobile station the burst data should be structured in a well-defined way. Therefore, the WCDMA contains a frame structure, which is divided into 15 slots, each of length 2/3 ms and thus the frame length is 10 ms (Figure 3.29).

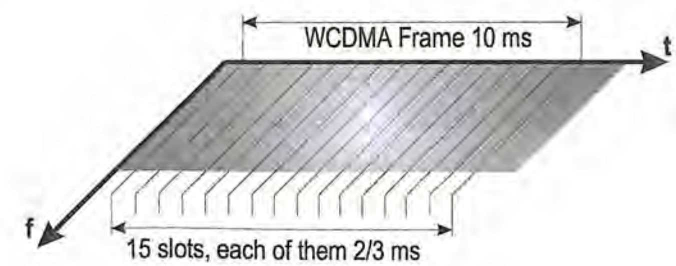


Figure 3.29. WCDMA frame

Based on this, one WCDMA frame is able to handle

$$\frac{0.010\ \text{s}}{0.00000026041\ \text{s}} \approx 38400\ \text{Chips}$$

One slot in the WCDMA frame contains

$$\frac{38400\ \text{Chips}}{15\ \text{Slots}} = 2560\ \text{Chips}$$

Unlike in GSM, WCDMA does not contain any super-, hyper- or multiframe structures. Instead, WCDMA frames are numbered by an SFN (System Frame Number). An SFN is used for internal synchronisation of UTRAN and timing of BCCH information transmission.

Figures 3.30 and 3.31 illustrate the uplink and downlink dedicated channel frame structures, respectively. As shown in the figures, the dedicated physical channels have different structure for uplink and downlink.

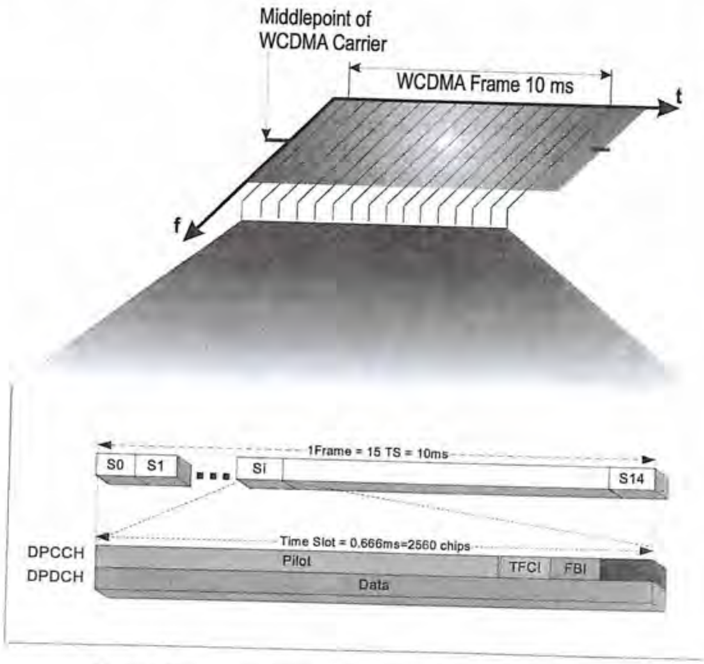


Figure 3.30. Frame structure for a UL dedicated physical channel

In the uplink case, the basic frame structure of the dedicated physical channel follows the downlink frame structure, but the main difference is that uplink dedicated channels cannot be

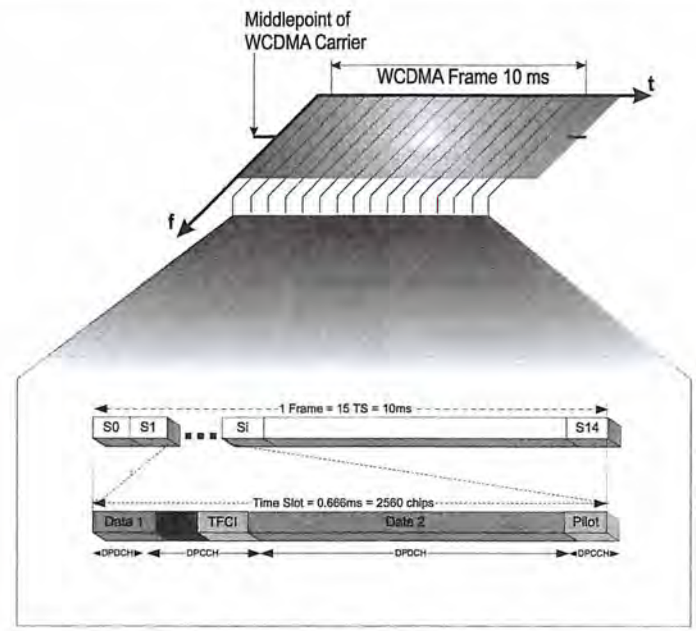


Figure 3.31. Frame structure of a DL dedicated physical channel

considered as a time multiplex of DPDCH and DPCCH. This leads to the fact that multicode operation is possible for the UL dedicated physical channels. It should be mentioned, however, when multicode is used, several parallel DPDCHs are transmitted using different channelisation codes with only one DPCCH per connection.

The uplink DPCCH consists of pilot bits to support channel estimation for coherent detection, Transmit Power Control (TPC) commands for transmit power adjusting, Feedback Information (FBI), and an optional Transport Format Combination Indicator (TFCI) to inform the receiver about the instantaneous parameters of the different transport channels multiplexed on the uplink DPDCH, and corresponds to the data transmitted in the same frame.

In the case of DL, every slot includes pilot bits, transmitted power control bits, a transport frame indicator, and data.

It needs to be mentioned that within one downlink DPCH, the dedicated data generated at layer 2 and above, i.e. the DCH, is transmitted in time multiplex with control information generated by layer 1 (known pilot bits, TPC commands, and an optional TFCI). Therefore, the downlink DPCH can be seen as a timed multiplex of downlink DPDCH and downlink DPCCH. In addition to the harmonisation compromise in standardisation, the reasons behind this approach are:

- To minimise the continuous transmission by mobile handset;
- To use the downlink orthogonal codes more efficiently;
- To minimise the power control delay by utilising the slot-offset between uplink and downlink slots.

For the common channels the structure is the same and the main difference between the common and dedicated channels is that in common channels the TPC bits are not used.

Finally, Table 3.4 summarises the main characteristics of the WCDMA-FDD scheme described previously. These technical parameters enable the WCDMA-FDD scheme to meet the system requirements for the 3G mobile system.

Table 3.4 WCDMA-FDD: main technical characteristics

Parameter	Specification
Multiple access	FDD: DS-CDMA
Duplex scheme	FDD
Chip rate (Mc/s)	3.84
Frame length (ms)	10
Channel coding	Convolutional coding ($R = 1/2, 1/3, 1/4, K = 9$); turbo code of $R = 1/2$, 1/3, 1/4 and $k = 4$
Interleaving	Inter/intraframe
Data modulation	FDD: DL:QPSK, UL dual channel QPSK
Spreading modulation	FDD: UL:BPSK, DL:QPSK
Power control	Closed loop (inner loop, and outer loop), open loop. Step size: 1–3 dB (UL); power cycle: 1500/s
Diversity	RAKE in both BS and MS; antenna diversity; transmit diversity
Inter-BS synchronisation	FDD: no accurate synchronisation needed
Detection	MS&BS: pilot symbol based coherent detection in UL, CPICH cannel estimation in DL
Multiuser detection	Supported (not at the first phases)
Service multiplexing	Variable mixed services per connection is supported
Multirate concept	Is supported by utilising a variable spreading factor and multicode
Handover	Intra-frequency soft and softer handovers are supported, inter-system and inter-frequency handovers are supported

4

UMTS Radio Access Network (UTRAN)

The main task of UTRAN is to create and maintain Radio Access Bearers (RAB) for communication between User Equipment (UE) and the Core Network (CN). With RAB the CN elements are given an illusion about a fixed communication path to the UE, thus releasing them from the need to take care of radio communication aspects. Referring to the network architecture models presented in Chapter 1, UTRAN realises certain parts of Quality of Service (QoS) architecture independently (Figure 4.1).

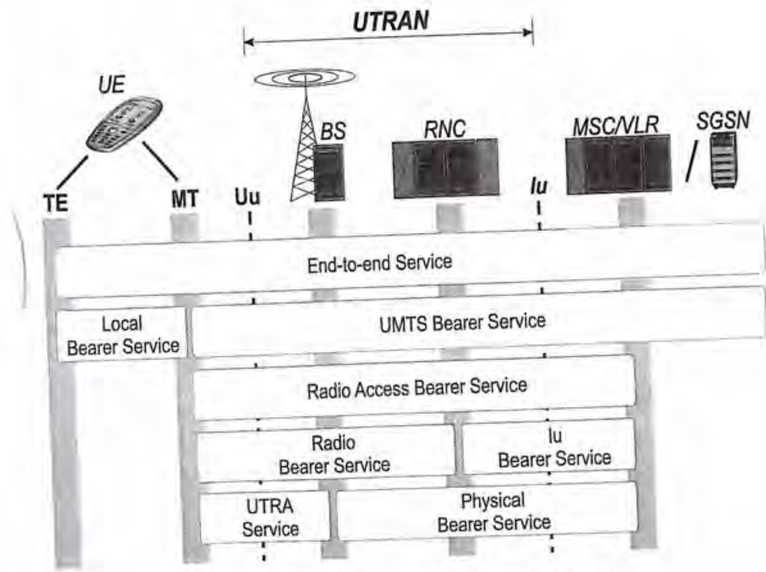


Figure 4.1. Bearer/QoS architecture in UTRAN

UTRAN is located between two open interfaces being Uu and Iu. From the bearer architecture point of view, the main task of UTRAN is to provide bearer service over these interfaces; in this respect the UTRAN controls Uu interface and in Iu interface the bearer service provision is done in co-operation with the CN.

The RAB fulfils the QoS requirements set by the CN. The handling of end-to-end QoS requirement in the CN and in the UE is the responsibility of Communication Management. Those requirements are then mapped onto the RAB, which is "visible" for the (Mobile Termination) MT and the CN. As said earlier, the main task of UTRAN is to create and maintain RAB so that the end-to-end QoS requirements are fulfilled in all respects.

One of the main ideas of this layered structure is to encapsulate the physical radio access, later it can be modified or replaced without changing the whole system. In addition, it is a known fact that the radio path is very complex and continually changing transmission media. This bearer architecture gives remarkable role to the Radio Network Controller (RNC), since the RNC and the CN map the end-to-end QoS requirements over the Iu interface and the RNC takes care of satisfying the QoS requirements over the radio path. These two bearers exist in the system because the Iu bearer is more stable in nature; the RB experiences more changes during the connection. For example, one UE may have three continuously changing RBs maintained between itself and the RNC, still the RNC has only one Iu bearer for this connection. This kind of situation occurs in context with soft handover, which is described in this chapter.

The physical basis of the end-to-end service in Uu interface is Universal Terrestrial Radio Access (UTRA) Service. As was explained in Chapter 3, UTRA Service is implemented with Wideband Code Division Multiple Access (WCDMA) radio technology and in the beginning of UMTS the used variant was WCDMA-FDD. From the QoS point of view the UTRA Service contains mechanisms that show how the end-to-end QoS requirements are mapped onto the physical radio path. Respectively, every Uu interface connection requires terrestrial counterpart through UTRAN. The equal terrestrial physical basis for the end-to-end service is Physical Bearer Service. This could be implemented with various technologies but in Third Generation Partnership Project (3GPP) R99 implementation the most probable alternative is Asynchronous Transfer Mode (ATM) over physical transmission media. Both ATM and some of the physical transmission alternatives are explained in Chapter 9. As UTRAN evolves the ATM will have Internet Protocol (IP) as an alternative way of implementation (refer to Chapter 2)

Referring to the conceptual model presented in Chapter 1, the protocols realising the RB Service belong to Access Stratum. The protocols above RB Service belong to Non-Access Stratum and are responsible for the UMTS bearer service.

4.1. UTRAN Architecture

Figure 4.2 illustrates the UTRAN architecture on the network element level. UTRAN consists of Radio Network Subsystems (RNS) and each RNS contains various amount of Base Stations (BS, or officially Node B) realising the Uu interface and one RNC.

The RNSs are separated from each other by UMTS Interface between RNCs (Iur) interface forming connections between two RNCs. The Iur, which has been specified as an open interface, carries both signalling and traffic information.

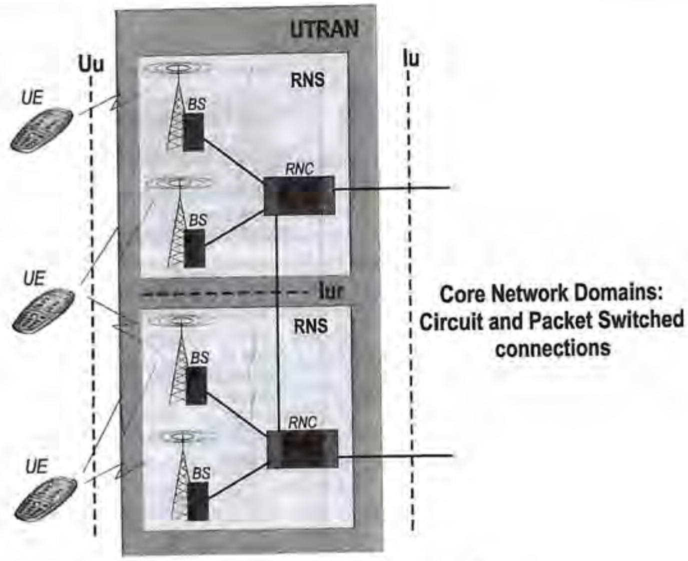


Figure 4.2. UTRAN architecture

4.2 Base Station

The Base Station (BS) is located between the Uu and UMTS Interface between RNC and BS (Iub) interfaces. Its main tasks are to establish the physical implementation of the Uu interface and, towards the network, the implementation of the Iub interface by utilising the protocol stacks specified for these interfaces. Realisation of the Uu interface means that the BS implements WCDMA radio access physical channels and transfers information from transport channels to the physical channels based on the arrangement determined by the RNC (for channels and their descriptions, refer to Chapter 3).

4.2.1 Base Station Structure

The internal structure of the BS is a vendor-dependent issue, but the logical structure, i.e. how a BS is treated within UTRAN, is generic. From the network point of view, the BS can be divided into several logical entities as shown in Figure 4.3.

On the Iub side, a BS is a collection of two entities, common transport and a number of Traffic Termination Points (TTP). The common transport represents those transport channels that are common for all UE in the cell and those used for initial access. The common transport entity also contains one Node B Control port used for operation and maintenance (O&M) purposes. One TTP consists of a number of *Node B Communication Contexts*. Node B Communication Context in turn consist of all dedicated resources required when the UE is

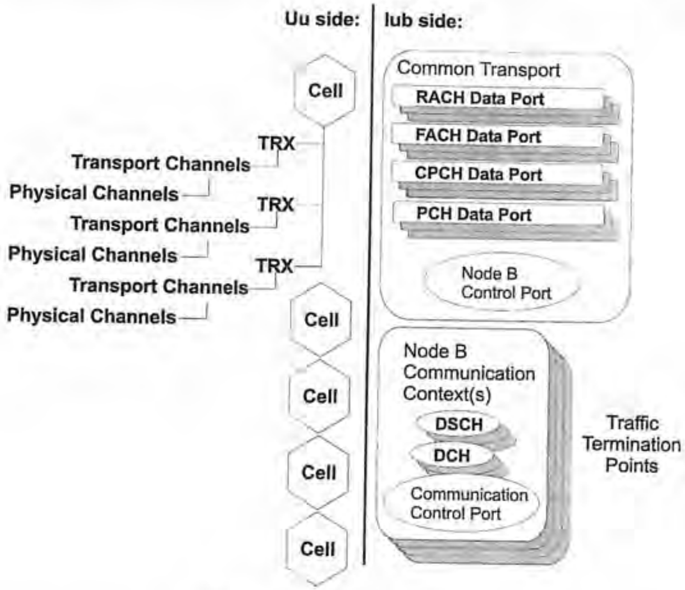


Figure 4.3. BS logical structure

in dedicated mode. Thus, one Node B Communication Context may contain at least one Dedicated Channel (DCH), for example. The exception is the Downlink Shared Channel (DSCH), which also belongs to Node B Communication Context. From the point of view of UMTS network infrastructures, the BS can be considered to be a logical O&M entity, which is subject to network management operations. In other words, this term describes physical BS and its circumstances, BS Site.

From the point of view of the radio network and its control, BS consists of several other logical entities called cells. A cell is the smallest radio network entity having its own identification number (cell ID), which is publicly visible for the UE. When the radio network is configured, it is actually the cell(s) data that is changed. The term sector stands for the physical occurrence of the cell, i.e. radio coverage.

Every cell has one scrambling code, UE recognises a cell by two values, scrambling code (when logging into a cell) and cell ID (for radio network topology). One cell may have several Transmitter-Receivers (TRXs, also called carriers) under it. The TRX of the cell delivers the broadcast information towards the UE. This is, the Primary Common Control Physical Channel (P-CCPCH) containing the Broadcast Channel (BCH) information is transmitted here. One TRX maintains physical channels through the Uu interface and these carry the transport channels containing actual information, which may be either common or dedicated in nature.

A cell may consist a minimum of one TRX. The TRX is physically a part of the BS performing various functions, data flows are converted from terrestrial Iub connection to the radio path and vice versa.

4.2.2 Modulation Method

From the system point of view, the used modulation method is of interest since it has a close relationship to the overall system capacity and performance. WCDMA uses Quadrature Phase Shift Keying (QPSK) as its modulation method in the downlink direction. To make it simple, this modulation method expresses a bit and its status with a different phase of the carrier. The bits in the modulation process are handled as pairs and thus, this generates four possible two-bit-combinations to be indicated.

In the modulation process, the data stream to be modulated, i.e. physical channels are first converted from serial to parallel format. After this conversion, the modulation process divides the data stream into two branches, I and Q. In the I branch the bit having value "1" represents + 180° phase shift and value "0" means the phase of the carrier has not shifted. In the Q branch the bit having value "1" represents + 90° phase shift and value "0" stands for − 90° phase shift. These branches, I and Q, are fed by the oscillator, I branch directly and Q branch with 90° shift. When the incoming data stream is combined with the oscillator output, the two-bit-combinations (one bit from I and one from Q branch) represent the phase shifting shown on the right-hand side of Figure 4.4. Thin lines in the square indicate the "paths" where the system can transfer itself from one state (phase) to another.

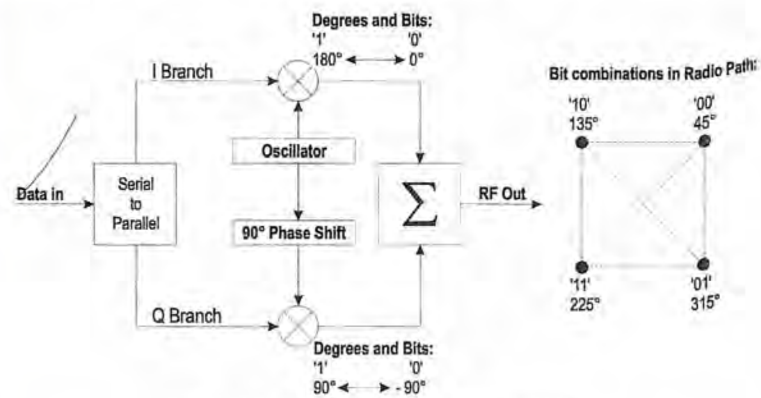


Figure 4.4. QPSK modulation process principle

This system works fine, but certain two-bit-combinations are more difficult to handle. For instance, when the bit values are changing from "00" to "11" it means 180° phase shifting which can be seen as very remarkable amplitude change. Very big amplitude changes cause problems especially if the bandwidth used in the radio connection is wide. In this case the BS must have linear amplifiers in order to guarantee that the amplitude changes are represented correctly throughout the used bandwidth.

To eliminate this too-rapidly-changing-amplitude problem, another variant of the QPSK is used: Offset Quadrature Phase Shift Keying (OQPSK). OQPSK introduces a time delay on the Q branch of the modulator and this time delay is equal to 0.5 bits (chips). This prevents 180° phase changes and limits them to 90° steps. Based on this, the transition from the combination "11" to "00" is now "11" – "10" – "00" and within the same time as in the QPSK (Figure 4.5).

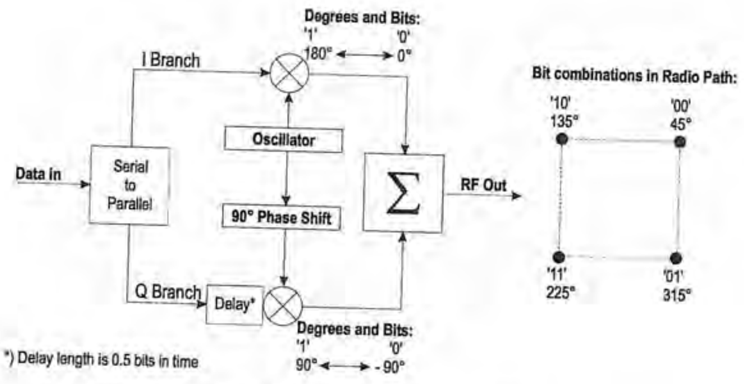


Figure 4.5. OQPSK modulation principle

The result is that the spectrum used for QPSK and OQPSK are the same but OQPSK has smoother signal. This allows the amplifiers to operate also on their non-linear operating area without problems. Based on these facts, the WCDMA actually uses both QPSK variants, the conventional QPSK in the downlink direction and the OQPSK in the uplink direction (Figure 4.6).

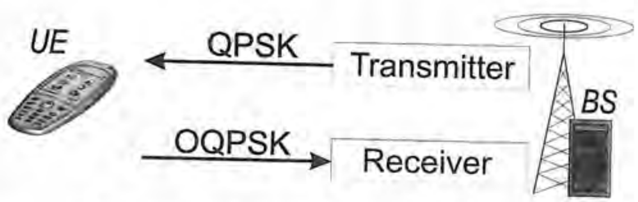


Figure 4.6. Used modulation methods and their directions

The conventional QPSK could be used in both directions but then the UE would suffer power consumption problems and high prices: the linear power amplifier should be very

accurate and thus very expensive. With OQPSK these amplifying problems could be avoided. From the BS point of view, the OQPSK is not "good enough" because the signal the BS transmits must be very accurate and the UE must achieve accurate synchronisation. With OQPSK, these features are compromised and thus the conventional QPSK is better for modulation performed on the transmission side of the BS.

4.2.3 Receiver Technique

The WCDMA utilises multipath propagation. This is, the transmitted signal propagates in many different ways from the transmitter to the receiver. On the other hand, to gain better capacity in the radio network, the transmit powers of the UE (and BSs) should be relatively small. This decreases interference in the radio interface and gives more space for other transmissions. This leads to a situation where it is very useful that both the UE and the BS are able to "collect" many weak level signals indicating the same transmission and to combine them together. This requires a special type of receiver. One example of this kind of arrangement is called RAKE. It should be noted that RAKE is not an acronym; it is real name for this type of receiver.

Therefore, the purpose of the RAKE receiver is to improve the received signal level by exploiting the multipath propagation characteristics of the radio wave, as signals propagating by different paths have different attenuation. A basic RAKE receiver consists of a number of fingers, a combiner, a matched filter, and a delay equaliser. Practically every finger can receive part of the transmitted signal, which can be either from the serving BS or from neighbouring BS. Different branches of the received signal are combined in such a way that the phase and amplitude deviations of the branches compensate for each other, resulting in a signal with remarkably higher signal strength than every individual signal branch. In order for the multipath signal branches to be distinguished from each other, there must be a specific delay range between consecutive branches.

In general, diversity techniques are efficient means to overcome the radio signal deterioration due to shadowing and fading which is discussed in Chapter 3. In addition, utilising diversity technique is a prerequisite for providing soft handover feature in cellular systems. Different diversity techniques may be used in mobile systems, that is, time diversity, space diversity, frequency diversity, etc. In WCDMA technology, typically polarisation diversity is utilised for both the uplink and downlink transmission. The purpose of multipath diversity is to resolve individual multipath components and combine them to obtain a sum signal component with better quality. Having used the RAKE receiver both in UE and the BS makes it possible to capture and combine different branches of the desired signal and improve the quality of the ultimate signal or data stream to be processed further. Maximum Ratio Combining (MRC) is the diversity algorithm used in the signal processing. The responsibility for diversity is distributed in radio access between UE, BS, and RNC depending on the link direction (uplink, downlink) and the position of the network element in the system architecture hierarchy.

4.2.4 Cell Capacity

In WCDMA technology, all the users share the common physical resource, frequency band in 5 MHz slices. All users of the WCDMA TRX co-exist on the frequency band at the same moment of time and different transactions are recognised with spreading codes. In the first UMTS systems, the UTRAN will use the WCDMA-FDD variant. In this variant, the Uu interface transmission directions use separate frequency bands. One of the most interesting questions (and one of the most confusing items) for people is the capacity of the WCDMA TRX. In Global System for Mobile communications (GSM) the TRX capacity calculation is a very straightforward procedure, but because in WCDMA the radio interface is handled differently and the system capacity is limited by variable factors, the capacity of the WCDMA TRX is not very easy to be determined.

The following text presents an idea of how to *roughly* estimate WCDMA TRX capacity theoretically and based on radio conditions. In order to simplify the issue, one must make some assumptions:

- All the subscribers under the TRX coverage area are equally distributed so that they have equal distances to the TRX antenna.
- The power level they use is the same and thus the interference they cause is on the same level.
- Subscribers under the TRX use the same base-band bit rate, i.e. also the same symbol rates.

Under these circumstances a value called Processing Gain (G_p) can be defined. Processing Gain is a relative indicator informing what is the relationship between the whole bandwidth available (B_{RF}) and the base-band bit rate ($B_{Information}$).

$$G_p = \frac{B_{RF}}{B_{Information}}$$

There is another way to express the G_p by using chip and data rates:

$$G_p = \frac{\text{Chiprate}}{\text{Datarate}}$$

Both ways (when announced in dB values) give as a result the improvement of the Signal to Noise Ratio (S/N) between the received signal and the output of the receiver.

Further on, the processing gain is actually the same as the spreading factor. It should be noted that, the base-band bit rate discussed here is the one achieved after the rate matching. In this process the original (user) bit rate is adjusted to the bearer bit rate. Bearer bit rates are somewhat fixed, e.g. 30 kb/s, 60 kb/s, 120 kb/s, 240 kb/s, 480 kb/s and 960 kb/s. The system chip rate is constant; 3.84 Mcps (3 840 000 chips/s). Hence, as an example the bearer having a bit rate 30 kb/s will have a spreading factor 128:

$$G_p = \frac{3\,840\,000}{30\,000} = 128 = \text{Spreading Factor}$$

The power P required for information transfer in one channel is a multiple of the energy used per bit and the base-band data rate.

$$P = E_b \times \text{Baseband Datarate}$$

On the other hand, it is known that the noise on one channel ($N_{Channel}$) using partially the whole bandwidth B_{RF} can be expressed as:

$$N_{Channel} = B_{RF} \times N_O$$

where N_O is the Noise Spectral Density (W/Hz)

Based on this, the signal to noise ratio is:

$$S/N = \frac{P}{N_{Channel}} = \frac{E_b \times \text{Baseband Datarate}}{B_{RF} \times N_O} = \frac{E_b/N_O}{G_p}$$

If assumed that there are X users under the TRX and the assumptions presented earlier in this chapter are applied, it means that there are (statistically thinking) $X - 1$ users causing interference to one user. This also indicates signal to noise ratio and when expressed in mathematical format the outcome is the following equation:

$$S/N = \frac{P}{P \times (X-1)} = \frac{1}{X-1}$$

If, further on, there are plenty of users (tens of them) then the equation could be simplified:

$$S/N = \frac{1}{X-1} \approx \frac{1}{X}$$

Now there are two different ways to calculate signal to noise ratio:

$$\frac{E_b/N_O}{G_p} \approx \frac{1}{X} \Rightarrow X \approx \frac{G_p}{E_b/N_O}$$

This equation is a *very rough* expression and should be used for estimation purposes only. The "official" way to calculate TRX capacity has several more parameters to be taken into account.

Example

Assume that the spreading factor used in the cell is 128 and for those transactions the E_b/N_O is 3 dB. How many users can the cell contain simultaneously assuming that there is 1 TRX available?

$$X \approx \frac{G_p}{E_b/N_O} = \frac{128}{3\text{ dB}} \approx \frac{128}{2} = 64 \text{ Users}$$

This is the maximum amount of users the TRX is able to handle, in theory when taking into account the intra-cell interference. In reality the neighbouring cells produce inter-cell interference. If assumed that the inter-cell interference is the same as intra-cell interference then the user amount is split to 32 (64/2 = 32).

The E_b/N_o relationship is the point of interest. is a constant-like numeral relationship which may have several values, which later are related to radio interface bearer bit rates. Thus it can be stated that the E_b/N_o relationship has a remarkable effect on the TRX/cell capacity as far as simultaneous user amount is concerned. In the E_b/N_o relationship, the following issues should be considered:

- N_o is a local constant value type, which also contains some receiver specific values.
- E_b is a changing value in nature and its dependencies are described in the following bullets.
- Processing gain/spreading factor: the bigger the spreading factor value, the smaller the E_b.

- The higher base-band bit rate used, the bigger the E_b will be (this is a direct consequence from the previous point).
- Distance between terminal and BS receiver: the longer distance, the bigger E_b.
- Terminal motion speed: the higher speed used, the bigger E_b.

The calculation above is a rough way to estimate TRX capacity, there are many other ways to do this. For further details, refer to the *WCDMA for UMTS* by Holma and Toskala (2001).

4.3 Radio Network Controller

The RNC is the switching and controlling element of the UTRAN. The RNC is located between the Iub and Iu interface. It also has the third interface called Iur for inter-RNS connections. The implementation of the RNC is vendor dependent but some generic points can be highlighted (Figure 4.7).

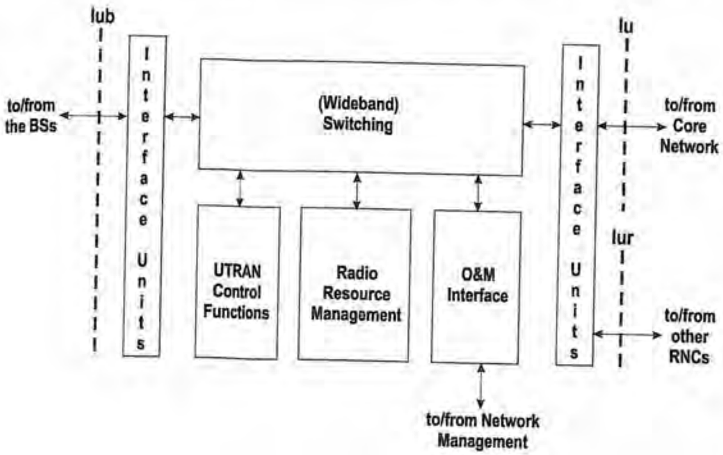


Figure 4.7. RNC logical structure

As explained earlier, the RNC sees the BS as two entities being common transport and collection of Node B Communication contexts. The RNC controlling these for a BS is called *Controlling RNC* (CRNC).

Referring to the bearers, the RNC is a switching point between the Iu bearer and RB(s). One radio connection between the UE and the RNC, carrying user data is a RB. The RB in turn is related to the UE Context. UE Context is a set of definitions required in Iub in order to arrange both common and dedicated connections between the UE and the RNC. Since UTRAN utilises macrodiversity, the UE may have several RBs between itself and the RNC. This situation is known as soft handover and is discussed later on in this chapter. The RNC holding the Iu bearer for certain UE is called *Serving RNC* (SRNC).

The third logical role the RNC may have is *Drifting RNC* (DRNC). When in this mode, the RNC allocates UE Context through itself, the request to perform this activity comes from the SRNC through the Iur interface.

Both SRNC and DRNC roles are functionalities, which may change their physical location. When the UE moves in the network performing soft handovers, the radio connection of the UE will be accessed entirely through a different RNC than the SRNC, which originally performed the first RB set-up for this UE. In this case the SRNC functionality will be transferred to the RNC, which practically handles the radio connection of the UE. This procedure is called SRNC or SRNS Relocation.

The whole functionality of RNC can be classified into two parts, UTRAN Radio Resource Management (RRM) and control functions. The RRM is a collection of algorithms used to guarantee the stability of the radio path and the QoS of radio connection by efficient sharing and managing of the radio resources. The UTRAN control functions include all of these functions related to set-up, maintenance and release of the RBs including the support functions for the RRM algorithms.

4.3.1 Radio Resource Management

As explained in Chapter 1, the RRM is a management responsibility solely taken care of by UTRAN. RRM is located in both UE and RNC inside UTRAN. RRM contains various algorithms, which aim to stabilise the radio path enabling it to fulfil the QoS criteria set by the service using the radio path (Figure 4.8).

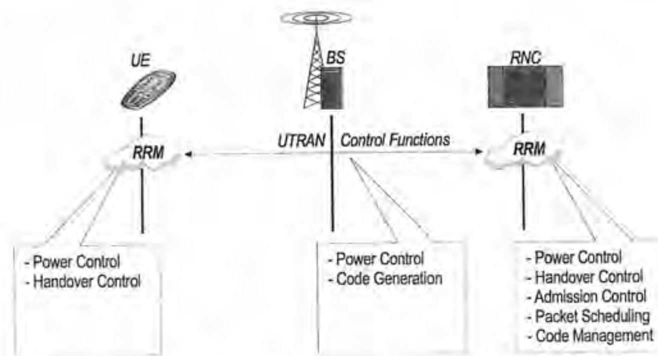


Figure 4.8. Radio Resource Management and Radio Resource Control

The RRM algorithms must deliver information over the radio path, which is named UTRA Service. The control protocol used for this purpose is the Radio Resource Control (RRC) protocol. The UTRAN control functionality's are discussed later in this chapter. The RRC

The RRM algorithms to be shortly presented here are:

- Handover Control
- Power Control
- Admission Control (AC) and Packet Scheduling
- Code Management

4.3.1.1 Handover Control

Handover is one of the essential means to guarantee the user mobility in a mobile communications network, where the subscriber can move around. Maintaining the traffic connection with a moving subscriber is made possible with the help of the *handover* function. The basic concept is simple: when the subscriber moves from the coverage area of one cell to another, a new connection with the target cell has to be set up and the connection with the old cell may be released. Controlling the handover mechanism is however, quite complicated issue in cellular systems and especially the Code Division Multiple Access (CDMA) system adds some interesting ingredients to it.

Reasons Behind the Handover

There are many reasons why handover procedures may be activated. The basic reason behind a handover is that the air interface connection does not fulfil the desired criteria set for it anymore and thus either the UE or the UTRAN initiates actions in order to improve the connection. In WCDMA, handover on-the-fly is used in context of circuit switched calls. In the case of packet switched calls, the handovers are mainly achieved when neither the network nor the UE has any packet transfer activity.

Regardless of the sort of handovers, they have the common nominator, handover criteria that is why the handover should be performed. The logic behind how the need for the handover is investigated is also quite common. The handover execution criteria depend mainly upon the handover strategy implemented in the system. However, most criteria behind the handover activating the rest on the signal quality, user mobility, traffic distribution, bandwidth, and so forth.

Signal Quality Reason handover occurs when the quality or the strength of the radio signal falls below certain parameters specified in the RNC. The deterioration of the signal is detected by the constant signal measurements carried out by both the UE and the BS. The signal quality reason handover may be applied both for the uplink and the downlink radio links.

Traffic reason handover occurs when the traffic capacity of a cell has reached its maximum or is approaching it. In such a case, the UE nears the edges of the cell with high load may be handed over to neighbouring cells with less traffic load. By using this sort of handover the system load can be distributed more uniformly and the needed coverage and capacity can be adapted efficiently to meet the demanded traffic within the network. Traffic reason handovers may rely on pre-emption or directed retry approaches.

The number of handovers is straightforwardly dependent on the degree of UE mobility. If we assume that the UE keeps on moving in the same direction then it can be said that the faster the UE is moving the more handovers it causes to the UTRAN. To avoid undesirable handovers the UE with high motion speed may be handed over, for instance, from micro cells to macro cells.

In case of the UE moving slowly or at all, on the other hand, it can be handed over from macro cells to micro cells to improve the radio signal strength and avoid consuming its battery.

The decision to perform a handover is always made by the RNC, that is currently serving the subscriber, except for the handover for traffic reasons. In the latter case the Mobile Switching Centre (MSC) may also make the decision. In addition to the above, many other reasons, for instance, change of services can also be a basis for the handover execution.

Handover Process

Figure 4.9 illustrates, a basic handover process consists of three main phases, including measurement phase, decision phase, and execution phase. The overall handover process discussed here is specifically related to the WCDMA system. Nevertheless, and as far as the basic principles are concerned, they are valid for any kind of cellular systems.

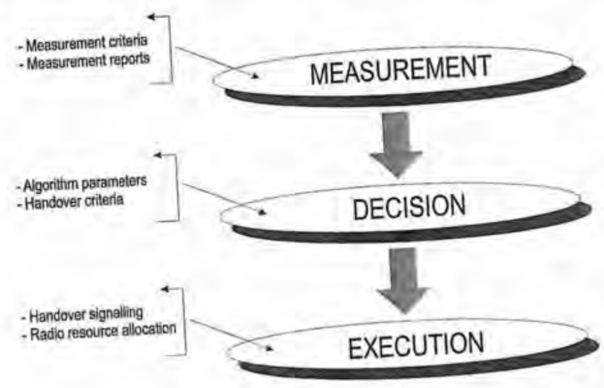


Figure 4.9. Handover process

Handover measurement provision is a pivotal task from the system performance standpoint, this is because of two reasons. Firstly, the signal strength of the radio channel may vary drastically due to the fading and signal path loss, resulting from the cell environment and user mobility. Secondly, an excess of measurement reports by UE or handover execution by network increases the overall signalling, which is not desired.

For the handover purposes and during the connection the UE continuously measures the signal strength concerning the neighbouring cells, and reports the results to the network, to the radio access controller like RNC in WCDMA system.

According to the 3GPP Technical Specification (TS) 25.331, the UE measurements may be grouped into different categories, depending on what the UE should measure. The different types of measurements are:

- Intra-frequency measurements include measurements on the strength of the downlink physical channels for signals with same frequencies.

- Inter-frequency measurements include measurements on the strength of the downlink physical channels for signals with different frequencies.
- Inter-system measurements covers measurements on the strength of the downlink physical channels belonging to another radio access system than UTRAN, for example GSM.
- Traffic volume measurements contain measurements of the uplink traffic volume.
- Quality measurements include measurements of quality parameters, for example the downlink transport block error rate.
- Internal measurements include measurements of UE transmission power and UE received signal level.

The measurement events, on the other hand, may be triggered based on the following criteria:

- Change of best cell
- Changes in the Primary Common Pilot Channel (CPICH) signal level
- Changes in the P-CCPCH signal level
- Changes in the Signal-to-Interference Ratio (SIR) level
- Changes in the Interference Signal Code Power (ISCP) level
- Periodical reporting
- Time-to-trigger

Therefore, the WCDMA specification provides various measurement criteria to support the handover mechanisms in the system. In order to improve the system performance it is pivotal to select the most appropriate measurement procedure and measurement criteria as well as filtering intervals in association with handover mechanisms. The handover signalling load can be optimised by fine-tuning trade-off between handover criteria, handover measurements, and the utilised traffic model in the network planning.

Decision phase consists of assessment of the overall QoS of the connection and comparing it with the requested QoS attributes and estimates from neighbouring cells. Depending on the outcome of this comparison, the handover procedure may or may not be triggered.

The SRNC checks whether the values indicated in the measurement reports trigger any criteria set. If they trigger, then it allows executing the handover.

As far as the handover decision-making is concerned, there are two main types of handover:

- Network Evaluated Handover (NEHO)
- Mobile Evaluated Handover (MEHO)

In the case of a NEHO procedure, the network SRNC makes the handover decision, while in the MEHO approach the UE mainly prepares the handover decision. In the case of combined NEHO and MEHO handover types, the decision is made jointly by the SRNC and the UE.

It should be noted that even in a MEHO handover the final decision about the handover execution is made by the SRNC. The reason is that the RNC is responsible for the overall RRM of the system, and thus it is aware of the system overall load and other necessary information needed for handover execution.

Handover decision-making is based on the measurements reported by the UE and the BS as well as the criteria set by the handover algorithm. The handover algorithms, as such, are not subject to standardisation; rather they are implementation-dependent aspects of the system.

Therefore, advanced handover algorithms may be utilised freely, based on the available parameters in association with the network element measurement capabilities, traffic distribution, network planning, network infrastructure, and overall traffic strategy applied by operators.

The general principles of a handover algorithm are presented in Figure 4.10. In this example it is assumed that the decision-making criteria of the algorithm are based on the pilot signal strength reported by the UE. The following terms and parameters are used in this handover algorithm example:

- *Upper threshold*: is the level at which the signal strength of the connection is at the maximum acceptable level in respect with the requested QoS.
- *Lower threshold*: is the level at which the signal strength of the connection is at the minimum acceptable level to satisfied the required QoS. Thus the signal strength of the connection should not fall below it.
- *Handover margin*: is a pre-defined parameter, which is set at the point where the signal strength of the neighbouring cell (B) has started to exceed the signal strength of current cell (A) by a certain amount and/or for a certain time.
- *Active Set*: is a list of signal branches (cells) through which the UE has simultaneously connection to the UTRAN.

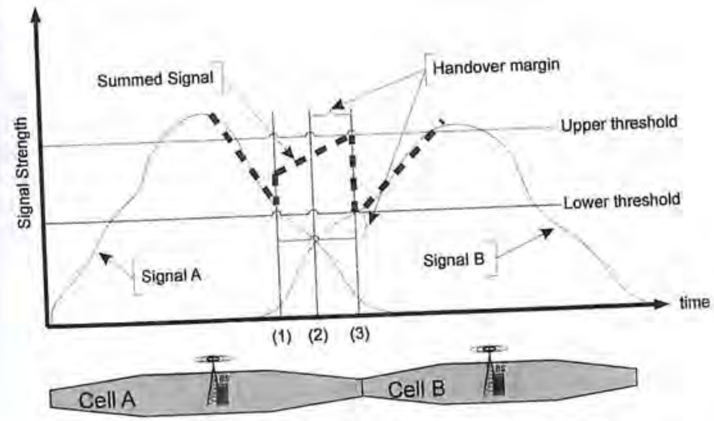


Figure 4.10. General principles of handover algorithms

Assume that a UE camping in cell A is moving towards cell B. As the UE is moving towards cell B the pilot signal A – to which the UE currently has a connection – deteriorates, approaching to the lower threshold as shown in Figure 4.10. This may result in handover triggering during the following steps which can be distinguished:

1. The strength of the signal A becomes equal to the defined lower threshold. On the other hand, based on the UE measurements the RNC recognises that there is already a neigh-

improving the quality of the connection. Therefore, it adds the signal B to the Active Set. Upon this event, the UE has two simultaneous connections to the UTRAN and hence it benefits from the summed signal, which consists of signal A and signal B.
2. At this point the quality of signal B starts to become better than signal A. Therefore, the RNC keeps this point as the starting point for the handover margin calculation.
3. The strength of signal B becomes equal or better than the defined lower threshold. Thus its strength is adequate to satisfy the required QoS of the connection. On the other hand, the strength of the summed signal exceeds the defined upper threshold, causing additional interference to the system. As a result, the RNC deletes signal A from the Active Set.

It should be noted that the size of the Active Set may vary but usually it ranges from 1 to 3 signals. In this example the size of Active Set is 2 between event (1) and (3).

Because the direction of UE motion varies randomly it is possible that it comes back towards the cell A instantly after the first handover. This results to a so-called *ping- pong* effect, which is harmful for the system in terms of capacity and overall performance. Using the handover margin or hysteresis parameter is to avoid the undesired handovers, which cause additional signalling load to the UTRAN.

Handover Types

Depending on the diversity used in association with handover mechanisms, they can be categorised as hard handover, soft handover, and softer handover. The hard handover can be further divided into intra-frequency and inter-frequency hard handovers. All of these mechanisms are provided by the UMTS system.

During the handover process, if the old connection is released before making the new connection it is called a *hard handover*. Therefore, there are not only lack of simultaneous

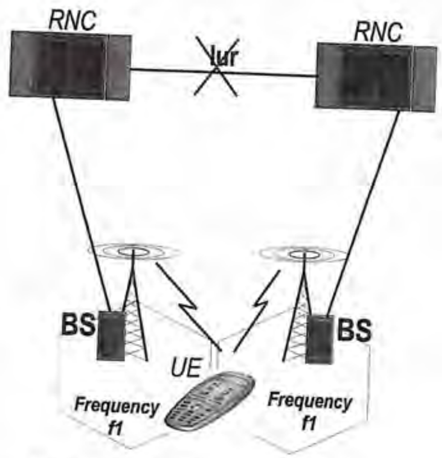


Figure 4.11. Intra-frequency hard handover

signals but also a very short cut in the connection, which is not distinguishable for the mobile user.

In case of *inter-frequency* hard handover the carrier frequency of the new radio access is different from the old carrier frequency to which the UE was connected. On the other hand, if the new carrier, to which the UE is accessed after the handover procedure is the same as the original carrier then there is an *intra-frequency* handover in question.

Figures 4.11 and 4.12 show hard handover situations when the neighbouring BSs may transmit with same frequencies or with the different frequency, respectively.

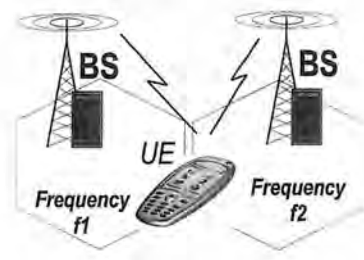


Figure 4.12. Inter-frequency hard handover

In Figure 4.11, the neighbouring RNC is not connected by Iur interface due to the radio network planning strategy or transmission reasons and hence the inter-RNC soft handover is not possible. Under these circumstances, intra-frequency hard handover is the only handover to support the seamless radio access connection and subscriber mobility from the old BS to the new BS. In fact, this leads to an inter RNC handover event, in which the MSC is also involved.

Generally, the frequency reuse factor is one for WCDMA, meaning that all BS's transmit on the same frequency and also all UE share a common frequency within the network. This does not mean, however, that the frequency reusing cannot be utilised in WCDMA at all. Therefore, if different carriers are allocated to the cells for some other reason, inter-frequency handover is required to ensure handover path from one cell to another cell in the cell cluster.

Inter-frequency handover also occurs in Hierarchical Cell Structure (HCS) network between separate cell layers, for instance, between macro cells and micro cells, which use different carrier frequencies within the same coverage area. In this case the inter-frequency handover is used not only because the UE would otherwise lose its connection to the network, but also in order to increase the system performance in terms of capacity and QoS. Inter-frequency hard handover is always a NEHO.

Moreover, inter-frequency handover may happen between two different radio access networks (RAN's), for instance, between GSM and WCDMA. In this context it can also be called inter-system handover (Figure 4.13). An inter-system handover is always a type of inter-frequency, since different frequencies are used in different systems.

The possibility to perform an inter-system handover is enabled in the WCDMA by a special functioning mode, *Compressed Mode (Slotted Mode)*. Compressed Mode is also

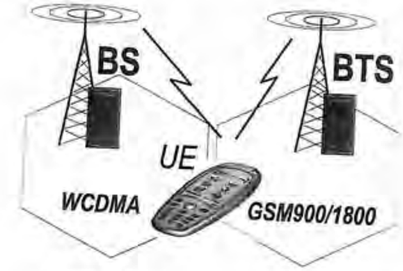


Figure 4.13. Inter-system hard handover

Slotted Mode, the spreading factor value of the channel can be reduced. As a consequence, the radio interface connection uses only part of the space in the WCDMA frame slot. The rest of the slot can be used for other purposes by the UE; for instance, the UE may measure surrounding GSM cells. In other words, this mechanism is the way to implement GSM/UMTS interoperability requirement in UTRAN. In addition, the Slotted Mode can be achieved by reducing the data rate using the higher layer controlling and also by reducing the symbol rate in association with the physical layer multiplexing. When the UE uses Uu interface in this mode, the contents of the WCDMA frame is "compressed" a bit in order to open a time window through which the UE is able to peek and decode the GSM Broadcast Control Channel (BCCH) information. Additionally, both the WCDMA RAN and GSM Base Station System (BSS) must be able to send each other's identity information on the BCCH's so that the UE is able to perform the decoding properly.

The inter-system WCDMA and GSM handover can be applied in areas where WCDMA and GSM systems co-exist. Inter-system handover is required to complement the coverage areas of each other in order to ensure continuity of services. The inter-system handover can also be used to control the load between GSM and WCDMA systems, when the coverage area

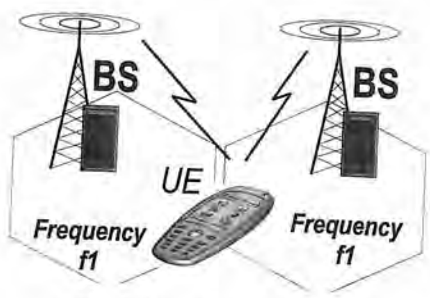


Figure 4.14. Intra-frequency soft handover

of the two systems overlap with each other. Other inter-system handover reasons might be the service requested by the UE and user's subscription profile.

Inter-system handover is a NEHO. However, also the UE must support inter-system handover completely. The RNC recognises the possibility of inter-system handover based on the configuration of the radio network mainly based on neighbour cell definitions and other control parameters. The same applies for the BSC in the GSM RAN side.

Unlike in hard handover, if a new connection is established before the old connection is released then the handover is called *soft handover*. In the WCDMA system, the majority of handovers are intra-frequency soft handovers. As is illustrated in Figure 4.14, in soft handover the neighbouring BS involved in the handover event transmits the same frequency.

Soft handover is performed between two cells belonging to different BS's but not necessarily to the same RNC. In any case the RNC involved in the soft handover must co-ordinate the execution of the soft handover over the Iur interface. In a soft handover event the source and target cells have the same frequency. In case of a circuit switched call, the terminal is actually performing soft handovers almost all the time if the radio network environment has small cells. There are several variations of soft handover, including softer and soft-softer handovers.

A softer handover is a handover by which a new signal is either added to or deleted from the Active Set, or replaced by a stronger signal within the different sectors, which are under the same BS (Figure 4.15).

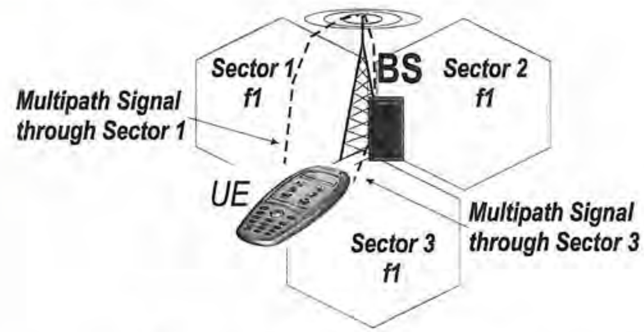


Figure 4.15. Intra-frequency softer handover

In softer handover, the BS transmits through one sector but receives from more than one sector. In this case the UE has active uplink radio connections with the network through more than one sector populating the same BS.

When soft and softer handovers occur simultaneously, the term soft-softer handover is usually used. A soft-softer handover may occur, for instance, in association with inter-RNC handover, while an inter-sector signal is added to the UE's Active Set along with adding a new signal via another cell controlled by another RNC.

Referring to the soft handover and Active Set, there are two terms describing the handling of the multipath components. These are microdiversity and macrodiversity.

Microdiversity means the situation where the propagating multipath components are combined in the BS as shown in Figure 4.16. The WCDMA utilises multipath propagation. This means that the BS RAKE receiver, which was already mentioned in 4.1.1.3, is able to determine, differentiate and sum up several signals received from the radio path. In reality, a signal sent to the radio path is reflected from the ground, water, buildings, etc. and at the receiving end the sent signal can be "seen" as many copies, all of them coming to the receiver at a slightly different phase and time. The microdiversity functionality at the BS level combines different signal paths received from one cell, and in the case of sectored BS, the outcome from different sectors, which is also referred to as softer handover.

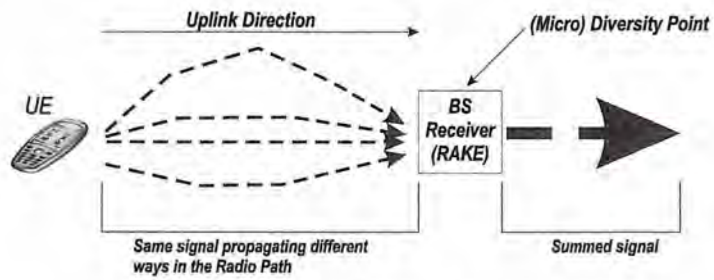


Figure 4.16. Micro diversity in BS

Because of the fact that the UE may use cells belonging to different BS's or even different RNC's the macrodiversity functionality also exists on the RNC level. However, the way of combining the signals is quite different than in the microdiversity case at BS because there is no RAKE receiver at RNC. Therefore, other approaches like the quality of data flow may be utilised to combine or select the desired data stream. Figure 4.17 presents a case in which the UE has a three-cell Active Set in use and one of those cells is connected to another RNC. In this case, the BS's firstly sum up the signal concerning the radio paths of their own and final summing of the data stream is done on the RNC level.

Concerning soft and softer handovers, the idea is that the subjective call quality will be better when the "final" signal is constructed from several sources (multipath). In GSM the subjective call quality depends on the transmission power used: it can be roughly stated that more power, the better quality. In WCDMA, the terminals cannot use a lot of power because transmission levels that are too high will start blocking the other users, thus, the better way to gain better subjective call quality is to utilise multipath propagation.

As a conclusion it can be stated that soft and softer handovers consume radio access capacity because the UE is occupying more than one radio link connection in the Uu interface. On the

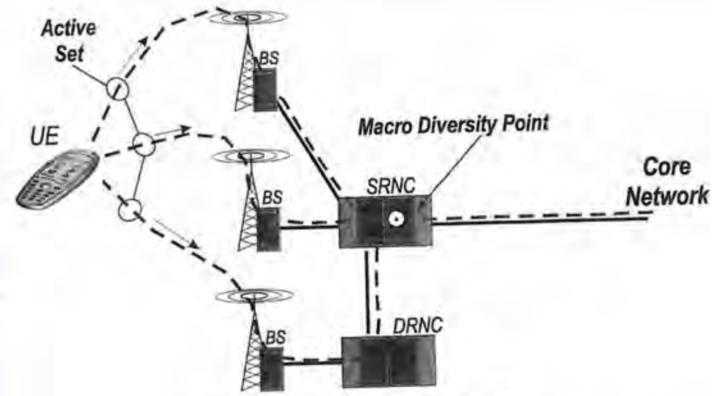


Figure 4.17. Macro diversity in RNC

other hand, the added capacity gained from the interference reduction is bigger and hence the system capacity is actually increased if soft and softer handovers are reasonably used.

From system structural architecture standpoint, UMTS network supports the following types of handovers:

- Intra BS/inter-cell handover (softer handover)
- Inter BS handover, including hard and soft handovers
- Inter RNC handover, including hard, soft, and soft-softer handovers
- Inter MSC handover
- Inter Serving GPRS Support Node (SGSN) handover
- Inter system handover

4.3.1.2 Power Control

Power control is an essential feature of any CDMA based cellular system. Without utilising an accurate power control mechanism, these systems do not operate. In the following subsections we first describe why power control is so essential for these cellular systems and what are the main factors behind this fact. We then describe what kind of power control mechanisms are utilised in the WCDMA-FDD radio access.

The main reasons for implementing power control are the near–far problem, interference dependent capacity of the WCDMA and the limited power source of the UE. Unlike Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA), which are bandwidth-limited multiple accesses, WCDMA is an interference-limited multiple access. In FDMA and TDMA, power control is applied to reduce inter-cell interference within the cellular system that arises from frequency reuse while in WCDMA systems, the purpose of the power control is mainly to reduce the intra-cell interference. Meeting these targets requires optimisation of the radio transmission power, meaning that the power of every transmitter is adjusted to the level required to meet the requested QoS. Determining the transmission power level is, however, a very sophisticated task due to dynamic variation of the radio channel.

Whatever the radio environment is, the received power should be at an acceptable level, for example at the BS for the uplink to support the requested QoS. The target of power control is to adjust the power to the desired level without any unnecessary increase in the UE transmit power. It takes care that, transmit power is just within the required level and neither higher nor less, taking into account the existing interference in the system.

The influence of multipath propagation characteristics and technical characteristics of WCDMA system, for instance, simultaneous bandwidth sharing and near–far phenomena result in the fact that power control, is essential for the WCDMA system, to overcome drawbacks caused by the radio environment and the nature of the electromagnetic wave. Without power control, phenomena like fading and interference, drive down the system stability and ultimately degrade its performance dramatically.

Maximising system capacity is an invaluable asset for both advanced cellular technology suppliers and cellular network operators. System capacity is maximised if the transmitted power of each terminal is controlled so that its signal arrives at the BS with the minimum required SIR. If a terminal's signal arrives at the BS with too low a received power value then the required QoS for the radio connection can not be met. If the received power value is too high, the performance of this terminal is good, however, interference to all the other terminal transmitters sharing the channel is increased and may result in unacceptable performance for other users, unless their number is reduced.

Due to the fact that in the WCDMA system the total bandwidth is shared simultaneously, other users can be experienced as a noise–like interference for a specific user. In case the power control mechanism is missing or operates imperfectly, common sharing of the bandwidth creates a severe problem, referred to as the near-far effect. In near-far situations the signal of the terminal that is close to the serving BS may dominate the signal of those terminals, which are far away from the same BS. Figure 4.18 illustrates the situation in

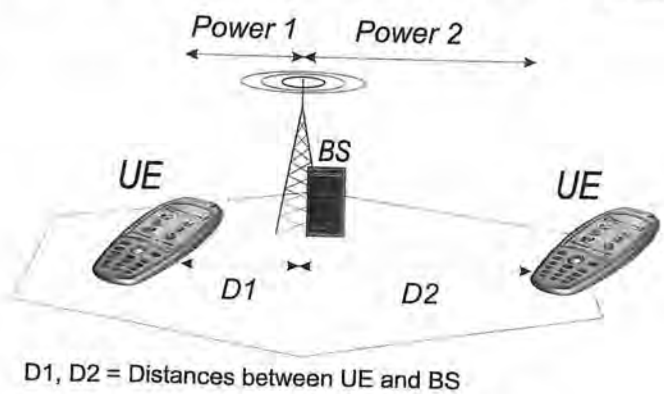


Figure 4.18. Near-far effect in the CDMA system

which the near-far problem may occur. The main factors that cause the near-far problem result from the path-loss variation of simultaneous users with different distances from the BS, fading variation, and other signal-power variation of the users caused by the radio wave propagation mechanisms described in Chapter 3.

In WCDMA, the near-far effect can be mitigated by applying power control mechanism, diversity techniques, soft handover, multi-user receiver and generally near-far resistance receivers. Because of the crucial drawback of the near-far effect on the performance of WCDMA system, its mitigating is one of the pivotal purposes of power control mechanisms. These mechanisms have considerable impact on the WCDMA system capacity.

Basic approaches to power control

Due to the facts described previously, it is relatively easy to determine that in the uplink case the optimal situation from the BS receiver point of view is that the power representing one UE's signal is always equal when compared to the other UE signals regardless of their distance from the BS. If so, the SIR will be optimal and the BS receiver is able to decode the maximal number of transmissions. In reality, however, the radio channel is extremely unstable and also the requested radio services vary for different users and even for the same user and during one radio connection. Therefore the transmission power of UE should be controlled very accurately by utilising efficient mechanisms.

To achieve this, power control has been well investigated and many power control algorithms have been developed since the origin of the CDMA scheme. These include distributed, centralised, synchronous, asynchronous, iterative, and non-iterative. Most of the existing algorithms utilise either SIR or transmit power as a reference point in the power control decision-making process.

The primary principle of Centralised Power Control (CPC) schemes is that they keep the overall power control mechanism centralised. As a result they require a central controller, which should have knowledge of all the radio connections in the RAN.

In contrary to CPC methods, distributed power control methods do not utilise central controller. Instead they distribute the controlling mechanism within the RAN and toward the edge of it. This feature makes them of special interest. The CPC approaches bring about added complexity, latency and network vulnerability. The main advantage of distributed power control algorithm is that it can respond more adaptively to variable QoS, which is vastly important for cellular systems with packet-based transmission characteristics like WCDMA.

Power Control Mechanism in UTRAN (WCDMA-FDD)

In WCDMA, power control is employed in both the uplink and the downlink. Downlink power control is basically for minimising the interference to other cells and compensating for other cells interference as well as achieving acceptable SIR. However, for downlink, power control is not as vital as it is for the up-link transmit power adjustment. It is still implemented also for the downlink, because it improves system performance by controlling interference to other cells.

The main target of the uplink power control is to mitigate the near-far problem by making the transmission power level received from all terminals as equal as possible at the home cell for the same QoS. Therefore uplink power control is for fine-tuning of terminal transmission power, resulting in the mitigation of the intra-cell interference and near-far effect. It should be mentioned that the power control mechanism specified for the WCDMA is, in principle, a distributed approach.

The power control mechanisms used in the GSM are clearly inadequate to guarantee this situation in WCDMA and thus the WCDMA has a different approach to the matter. In GSM, the power control is applied for the connection once or twice per second, but due to its critical nature in WCDMA, the power used in the connection is adjusted 1500 times per second, that is the power control cycle is repeated in association with each radio frame in association with DCH. Therefore, the power adjustment steps are considerably faster than in GSM.

To manage the power control properly in WCDMA, the system uses two different defined power control mechanisms as shown in Figure 4.19. These power control mechanisms are:

- Open Loop Power Control (OLPC)
- Closed Loop Power Control (CLPC), including inner and outer Loop Power Control mechanisms

By applying all of these different power control mechanisms together, the UTRAN benefits from the advantages of CPC as well by overlaying the inner closed loop control with the outer closed loop control mechanism in order to keep the target SIR in an acceptable level.

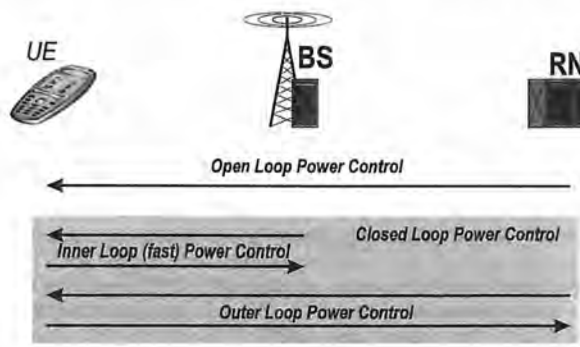


Figure 4.19. WCDMA power control mechanisms

Open Loop Power Control (OLPC)

In the OLPC, which is basically used for uplink power adjusting, the UE adjusts its transmission power based on an estimate of the received signal level from the BS CPICH when the UE is in idle mode and prior to Physical Random Access Channel (PRACH) transmission. In addition to that, the UE receives information about the allowed power parameters from the cell BCCH when in idle mode. The UE evaluates the path loss occurring and based on this difference together with figures received from the BCCH and the UE it is able to estimate what might be an appropriate power level to initialise the connection.

Figure 4.20 illustrates the OLPC as applied for the uplink case. In this process, the UE estimates the transmission signal strength by measuring the received power level of the pilot

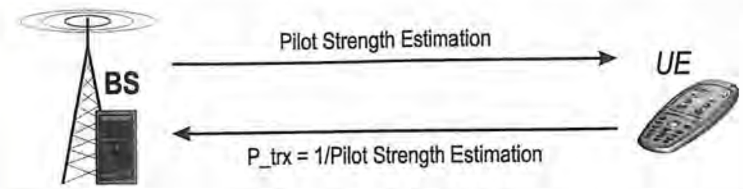


Figure 4.20. OLPC for uplink

signal from the BS in the downlink, and adjusts its transmission power level in a way that is inversely proportional to the pilot signal power level. Consequently, the stronger the received pilot signal, the lower the UE transmitted power.

In the case of Frequency Division Duplex (FDD)-based WCDMA, OLPC alone is not adequate for adjusting the UE transmission power, because fading characteristics of the radio channel vary rapidly and independently for the uplink and the downlink. Therefore, in order to compensate the rapid changes in the signal strength CLPC mechanism is also needed. Nevertheless, OLPC is useful for determining the initial value of transmitted power and mitigating drawbacks of the log-norm-distributed path-loss and shadowing.

Closed Loop Power Control (CLPC)

CLPC is utilised for adjusting the transmission power when the radio connection has already been established. Its main target is to compensate the effect of rapid changes in the radio signal strength and hence it should be fast enough to respond to those changes.

Figure 4.21 illustrates the basic uplink CLPC mechanism specified for WCDMA. In this case the BS commands the UE to either increase or decrease its transmission power with a cycle of 1.5 kHz (1500 times per second) by 1, 2, or 3 dB step-sizes. The decision whether to increase or decrease the power is based on the received SIR estimated by the BS. When the BS receives the UE signal it compares the signal strength with the pre-defined threshold value

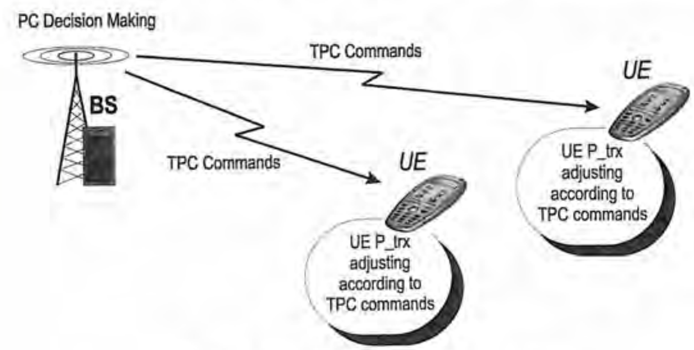


Figure 4.21. Basic uplink CLPC mechanism

at the BS. If the UE transmission power exceeds the threshold value, the BS sends a Transmission Power Command (TPC) to the UE to decrease its signal power. If the received signal is lower than the threshold target the BS sends a command to the UE to increase its transmission power. It should be emphasised that various measurement parameters, for example SIR, signal strength, Frame Error Ratio (FER) and Bit Error Ratio (BER) can be used for comparing the quality of the received power and making a decision for controlling transmission power.

It should be also noted that CLPC is utilised for adjusting transmission power in the downlink as well. In the case of the downlink CLPC, the roles of the BS and the UE are interchanged. That is, the UE compares the received signal strength from the BS with a predefined threshold and sends the TPC to the BS to adjust its transmission power accordingly.

In WCDMA, CLPC mechanism consists of Inner Loop and Outer Loop variants. What was described so far, was related to the Inner Loop variant, which is the fastest loop in WCDMA power control mechanism and hence it is occasionally referred to as fast power control.

Another variant of the CLPC is the Outer Loop Power Control (OLPC) mechanism. The main target of OLPC is to keep the target SIR for the uplink Inner Loop Power Control mechanism in an appreciated quality level. Thanks to the macrodiversity, the RNC is aware of the current radio connection conditions and quality. Therefore, the RNC is able to define the allowed power levels of the cell and target SIR to be used by the BS when determining the TPCs. In order to maintain the quality of the radio connection, the RNC uses this power control method to adjust the target SIR and keep the variation of the quality of the connection in control. By doing this, the network is able to compensate changes in the radio interface propagation conditions and to achieve the maximum target quality for the connection BER and the FER observation. In fact, the OLPC fine-tunes the performance of the Inner Loop Power Control.

Together, the Open and CLPC mechanisms have considerable impact on in the terminal's battery-life and overall system capacity in any cellular system and specifically in CDMA based mobile systems.

Power Control in Specific Cases

In addition to ordinary power control mechanisms used in WCDMA, there are additional approaches to cope with specific cases. These include controlling transmission power associated with soft handover, Site Selection Diversity (SSDT), and compressed mode (Slotted Mode).

In soft handover state, the transmission power of the UE is adjusted based on the selection of the most suitable power control command from those TPC's that it receives from different BS's to which it has simultaneous radio links (Figure 4.22). In this case because the UE receives more than one TPC command from the different BS's independently, the received TPC commands may differ from each other. This may result from the fact that power control commands are not efficiently protected against errors or simply it may result from the network environment. This leads to a conflict situation for the UE. The basic approach to resolve this problem is that, if at least one of the TPC commands refers to the decrease in transmission power then the UE decreases its power. It is also defined that the UE can utilise a threshold for detecting the reliable commands, based on this it can decide whether to increase or decrease its transmission power.

SSDT is another special power control solution. The principle of SSDT is that the BS with

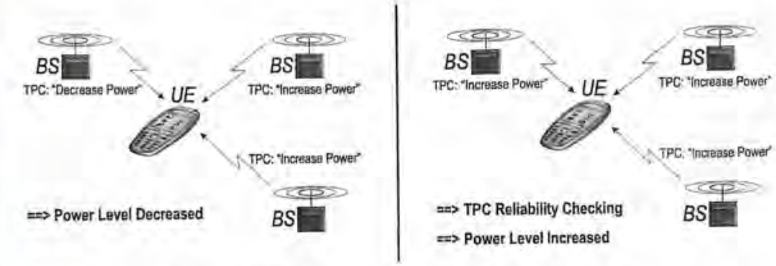


Figure 4.22. Power control associated to soft handover

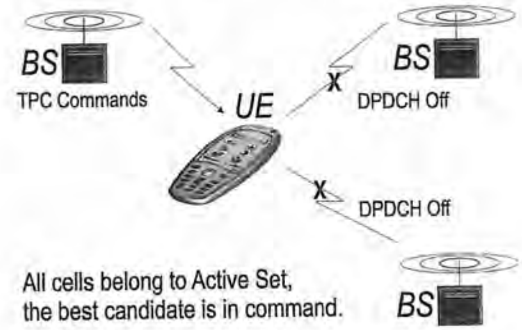


Figure 4.23. Site Selection Diversity (SSDT)

the strongest signal is dynamically chosen as the only transmitting BS (Figure 4.23). Then, the other BS's, to which the UE has simultaneously radio connectioned turn off their Dedicated Physical Data Channels (DPDCH). Therefore, the transmit power is adjusted based on the power control commands of the BS with the strongest signal. It seems that this method reduces the downlink interference generated while the UE is in soft handover.

Yet in compressed mode, the transmission and reception of the BS and UE are ceased for a predefined period to make room for performing the inter frequency radio measurements, for instance in conjunction with inter system handover event. As a result there is also a break in the transmission power adjusting mechanism. In this case the receiver of the power control commands, for instance, the UE for the uplink case is allowed to increase or decrease it's transmit power with larger step-sizes to reach the desirable SIR level as fast as possible.

4.3.1.3 Admission Control and Packet Scheduler

WCDMA radio access has several limiting factors, some of them being absolute and others environment dependent. The most important, and at the same time the most difficult to control is the interference occurring in the radio path. Due to the nature and basic characteristics of WCDMA, every UE accessing the network generates a signal and simultaneously this signal can be interpreted to be interference, from the other UE point of view. When the WCDMA cellular network is planned, one of the basic criteria for planning is to define the acceptable interference level with which the network is expected to function correctly. This planning-based value and the actual signals the UE transmit set practical limits for the radio interface capacity.

To be more specific, a defined SIR value is used in this context. Based on radio network planning the network is, in theory stable, as long as the defined SIR level is not exceeded within the cell. Practically, it means that in the BS receiver, the interference and the signal must have a certain level of power difference in order to extract one signal (code) out from the other signals using the same carrier. If the power distance between interfering components and the signal is too small the BS is unable to extract an individual signal (code) out from the carrier any more. Every UE having a bearer active through the cell "consumes" a part of the SIR and the cell is used up to its maximum level when the BS receiver is unable to extract the signal(s) from the carrier.

The main task of AC is to estimate whether a new call can have access to the system without sacrificing the bearer requirements of existing calls (Figure 4.24). Thus the AC algorithm should predict the load of the cell if the new call is admitted. It should be noted that the availability of the terrestrial transmission resources is verified, too. Based on the AC, the RNC either grants or rejects access.

Figure 4.24. Admission Control (AC)

Theoretically, it can be found out that the SIR or Interference Margin has direct relationship with the Cell Load. If we express the Cell Load with Load Factor as a parameter, which expresses the cell percentual load as shown in Figure 4.25 and mark the Interference Margin with I, it leads to the following equation:

$$I = 10 \times \text{Log}\left(\frac{1}{1 - \text{Load Factor}}\right)$$

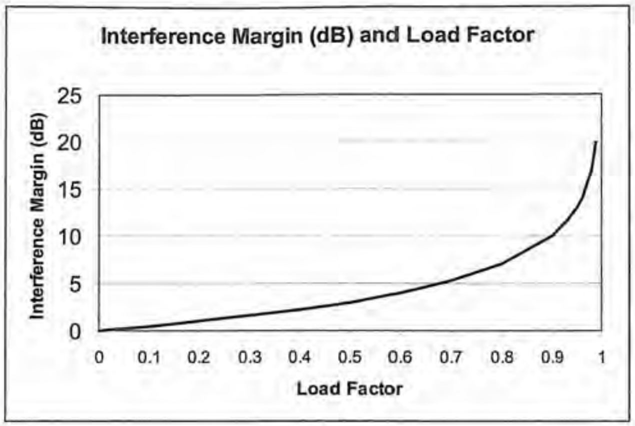


Figure 4.25. Interference margin as function of cell load

When placing together the Interference Margins calculated with different Load Factor values we arrive at the results shown in Figure 4.25.

Based on the graph it is fairly easy to indicate that when the Cell Load exceeds 70%, the interference in that cell will be very difficult to control. This is why the WCDMA radio network is normally dimensioned with expected capacity equivalent to Load Factor value 0.5 (50%), this value has a safety margin in it and the network will most likely operate steadily.

Due to the nature of traffic in the UMTS system both Real Time (RT) and Non-Real Time (NRT) traffic should be controlled and balanced carefully. The ways of controlling the RT and NRT traffic are distinctly different. The main difference is that for a RT (circuit call) service a RAB is maintained continuously in a dedicated state. In case of a NRT service (packet connection), the channel state varies as the consequence of the RB bit rate variation as well as the system load. So, the bit rate variation and the bursting characteristic of packet connection should be taken into account during packet connection admission process.

Figure 4.26 illustrates the main characteristics of a packet connection based on a World Wide Web (WWW) browsing session. During a WWW packet connection the user typically sends a WWW address to the network for data fetching purposes. As a response, the network downloads data, typically an HyperText Markup Language (HTML)–page , from the desired address. The desired address may also lead to document/file download. At the next step, a normal user is consuming a certain amount of time for studying the information, referring to it as the reading time. During the reading time there is no need for the AC to keep the RAB in a dedicated state, but rather the radio resources should be used for other purposes, e.g. other incoming and outgoing circuit or packet switched connections. The overall packet connection information (session) however, is maintained in upper level. So, if after the reading time the user wants to have another page downloaded, the pre-condition for the call is already available.

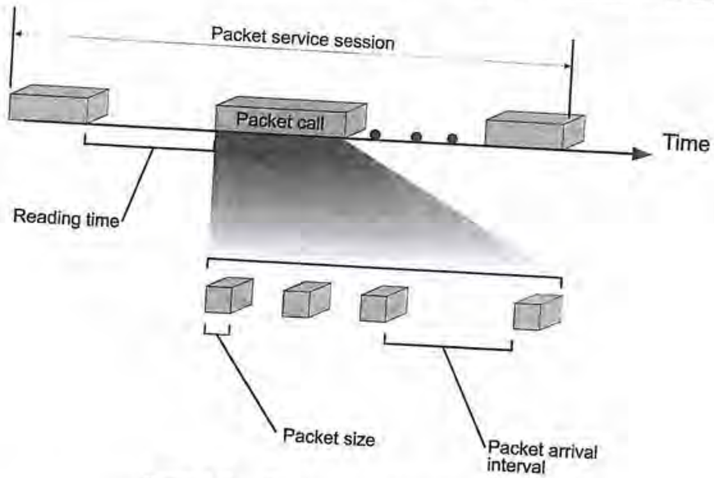


Figure 4.26. Characteristic of a packet service session.

The AC is responsible for the handling of packet connections with bursty traffic, having a very random arrival time, number of packet call per session, reading time, as well as number of packets within a call. Therefore, the AC must utilise very sophisticated traffic models and statistical approaches in order to control optimally the requested RAB(s). By scheduling the NRT RABs, and accepting, queuing, or rejecting the RT RABs. It is the responsibility of the AC to keep and control the QoS of the accepted RABs and their influence to the overall performance of the UTRAN.

4.3.1.4 Code Management

Both channelisation and scrambling codes used in the Uu interface connections are managed by the RNC. In principle, the BS could manage them, but then the system may behave unstable when the RNC is otherwise controlling the radio resources in terms of, for example soft handovers. When the codes are managed by the RNC, it is also easier to allocate Iub data ports for multipath connections.

The Uu interface requires two kinds of codes for proper functionality. Part of the codes must correlate with each other to a certain extent. The others must be orthogonal or they do not correlate at all. Every cell uses one scrambling code; the UE is able to make separation between cells by recognising this code. Under every scrambling code the RNC has a set of channelisation codes. This set is the same under every scrambling code. The BCH information is coded with a scrambling code value and thus the UE must find the correct scrambling code value first in order to access the cell. When a connection between the UE and the network is established, the channels used must be separated. The channelisation codes are used for this purpose. The information sent over the Uu interface is spread with a spreading code per channel and the spreading code used is scrambling code × channelisation code.

4.3.2 UTRAN Control Functions

In order for the UTRAN to control and manage the RBs, which is essential to provide the RAB service, it should perform other functions in addition to the RRM algorithms. These can be classified as:

- System Information Broadcasting
- Random Access and Signalling Bearer Set-up
- RB Management
- UTRAN Security Functions
- UTRAN level Mobility Management (MM)
- Database Handling
- UE positioning

4.3.2.1 System Information Broadcasting

An important function of the RNC is to handle the system information task. System information is used to maintain both the radio connections between the UE and the UTRAN and also control the overall operation of the UTRAN. The RNC broadcasts the system information elements to assist the UTRAN controlling functions by providing the UE with the essential data needed when communicating with the UTRAN. For example, for providing radio measurement criteria, paging occasion indication, radio path information, assistance data for positioning purposes, etc. The system information can be received by the UE both in idle mode and all connected states when it has been identified by the UTRAN. System information services may also be used, for example by the CN to provide broadcast services. The RNC utilises point-to-multi-point system information broadcasting to keep the UE in touch with the UTRAN when necessary.

From protocol architecture point of view, the system information broadcast functionality is part of the RRC and it is terminated at the RNC.

Figure 4.27 illustrates the system information structure, which together with segmentation class information is used by the RNC as basis for controlling the system information segmentation and scheduling. As shown, the system information elements are organised based on the Master Information Block (MIB), two optional Scheduling Blocks (SB) and System Information Blocks (SIB), which contains actual system information. 3GPP specification TS 25.331 defines up to 17 different types of SIB. Some of those may also contain sub-SIBs. A MIB contains reference and scheduling information to a number of SIBs in a cell. It may also contain reference and scheduling information to one or two SBs, which in turn gives reference and scheduling information for additional SIBs. Therefore, only the MIB and SBs can contain scheduling information for a SIB.

Because SIBs are characterised differently and grouped-specifically in terms of their repetition rate and criticality, the RNC can use them for different purposes, for example it uses SIB#1 to inform the UE about timers and counters to be used in idle and connected mode. Other examples are SIB#2 and SIB#3 which the RNC uses to inform the UE about the UTRAN level MM and cell selection and re-selection, respectively.

The RNC may handle the constructing process together with the BS, for example in modifying the system information scheduling. Nevertheless the controlling function of

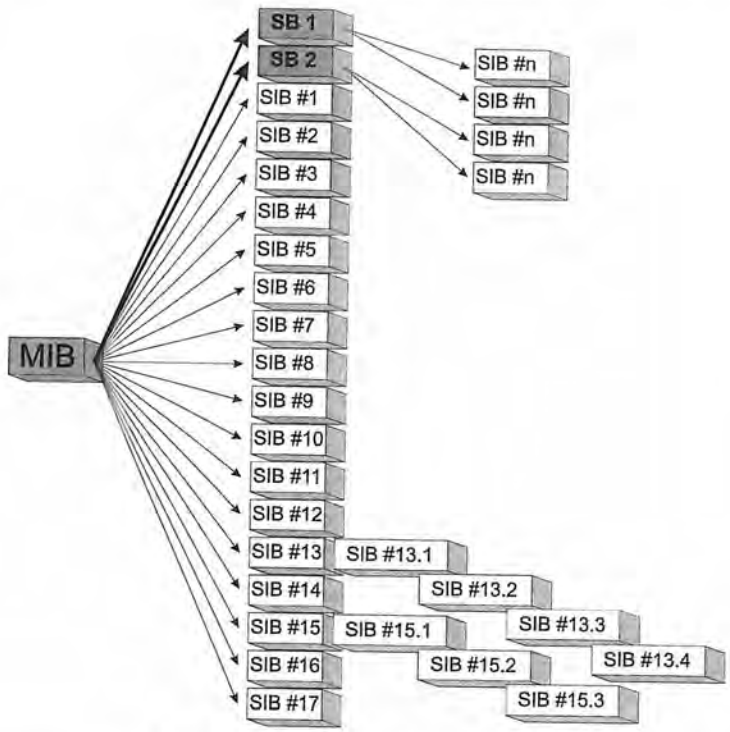


Figure 4.27. Basic structure of system information

system information belongs entirely to the RNC. Once the RNC has structured the system information data it forwards the information to the BS to be broadcasted over air interface to the UE. Based on the air interface situation the BS informs the RNC about the ability to broadcast the system information on the Uu interface (Figure 4.28).

4.3.2.2 Initial Access and Signalling Connection Management

Before the RNC can map any requested RABs to the RB, it needs to create the signalling connection between the UE and the CN. In this context, 3GPP specification Technical Report (TR) 25.990 defines the *signalling connection* as an acknowledged-mode link between the UE and the CN to transfer higher layer information between the entities in the Non Access Stratum (NAS). In order to do this the RNC uses the RRC connection services in creating the *Signalling Radio Bearer* (SRB) between the UE and the UTRAN to provide the transferring services for the signalling connection. Once the RNC has created the SRB, it uses the first

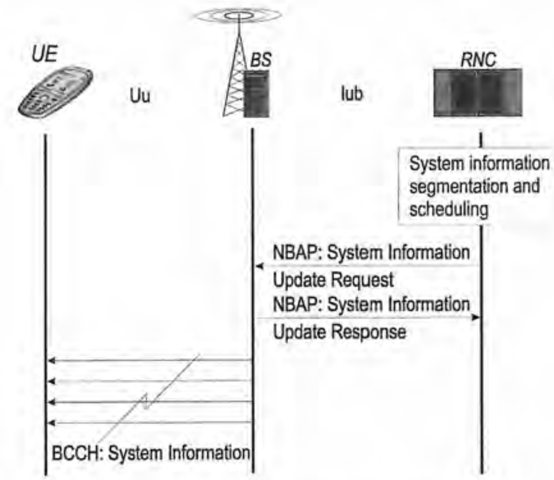


Figure 4.28. System information broadcasting

RLC RBs to convert and transfer the signalling connection through the UTRAN to the UE. *RB* is defined as the services provided by the RLC layer for transfer of user data between the UE and the RNC. It is however, defined that the primary RBs are also used for signalling connection purposes (Figure 4.29).

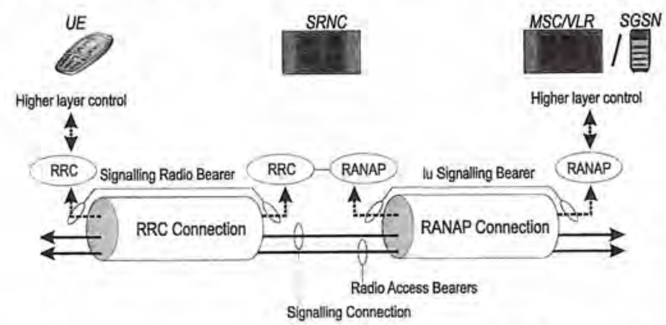


Figure 4.29. Relation between the SRB, RB, signalling connection, RAB and the role of RNC

The whole process starts when the UE enters the idle mode by turning on its power. After power-on, the UE attempts to make contact with the UTRAN. The UE looks for a suitable cell in the UTRAN and chooses the cell to provide available services, and tunes to its control channel. This choosing is known as "camping on the cell". The camping includes a cell search, where the UE searches for a cell and determines the downlink scrambling code and frame synchronisation of that cell. The cell search is down within different steps.

During the first step the UE uses the Synchronisation Channel (SCH)'s primary synchronisation code, which is common to all cells, to acquire slot synchronisation to a cell. During the second step, the UE uses the SCH's secondary synchronisation code to find frame synchronisation and identify the code group of the cell found in the first step. During the third and last step, the UE determines the exact primary scrambling code used by the cell found within the UTRAN. If the UE identifies the primary scrambling code then the P-CCPCH can be detected and the system and cell specific broadcast information defined by the RNC can be read.

In this regard the UE should be aware of the way the system information is structured and scheduled by the RNC based on the MIBs, SBs and SIBs. If the Public Land Mobile Network (PLMN) Identity found from the MIB on BCH matches the PLMN (or list of PLMNs) that the UE is searching for, it can continue to read the remaining BCH information, for example parameters for the configuration of the common physical channels in the cell (including PRACH and secondary CCPCH). Otherwise the UE may store the identity of the found PLMN for possible future use and restarts the cell search mechanism.

The first cell search for a PLMN is normally hardest to the UE, since it has to scan through a number of scrambling codes when trying to find the correct one. Once the UE obtains the necessary information to capture the BS controlled by the corresponding RNC it can request the *initial access* to the UTRAN, resulting in the transition from idle mode to the connected mode.

Until now the RNC was involved in the controlling function in terms of system information broadcasting and controlling the related BS to which UE was trying to get radio connection. Thereupon the RNC participates actively in controlling the access provisioning by considering the context of RRC connection set-up message requested by the UE via the Random Access Channel (RACH). In this association, the RNC has a central role to control the radio connection by checking the UE identity, the reason for the requested RRC connection and the UE capability, which helps it in allocating the initial signalling RB for the UE. Based on this information the RNC decides whether or not to allocate the signalling RB to the UE, which is used to carry the rest of the signalling for initial access and providing all the services after that. In case the RNC does not accept the access request then the UE can restart the initial access within a predefined time span.

No matter from which direction the higher layer service is requested, as is shown in the Figure 4.30, the RRC connection requested for creating the SRB is always started by the UE. When the RNC receives the RRC connection set-up request, it sets up a radio link over Iub interface to the BS to which the UE is going to have radio connection. In this context, the *Radio Link* is defined as a logical association between a single UE and a single UTRAN access point and its physical realisation comprises one or more RB transmissions. If this step is successful then the RNC informs the UE that the RRC connection is set-up and the UE responses to the RNC concerning the RRC connection completion.

Now the SRB is ready between the RNC and the UE and hence the RNC can convert and transfer the signalling connections and RABs between the UE and the CN. It should be mentioned that regardless of how many signalling connections and RABs exist between the UE and the CN, there is only one RRC connection used by the RNC to control and

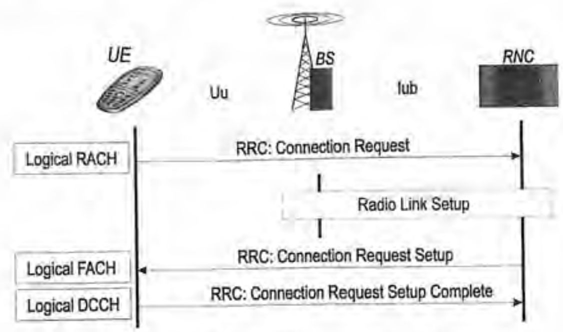


Figure 4.30. RRC connection set-up

transfer them between the UE and UTRAN. It is the function of RNC to reconfigure the lower layer services needed, based on the amount of the signalling connections and RABs variation, which should be transferred.

4.3.2.3 Radio Bearer Management

Once the signalling Radio Bearer (RB) is established between the radio network and the UE as was previously described, it is time for the RNC to map the requested RAB(s) onto the RB for transferring between the UTRAN and the UE. In this way the RNC creates the illusion about a fixed bearer between the CN and the UE.

After the requested RABs have been negotiated between the UE and CN via the signalling connection it is forwarded to the RNC for further actions. In this regard the main function of the RNC is to analyse the attributes of the requested RAB(s), evaluate the radio resources needed for these, activate or reconfigure the radio channels by using the lower layer services and map the requested RAB(s) to the RB. Therefore, the RNC covers all of the control functions related to the supporting of the RAB(s) as well as configuring the available radio resources for that purpose.

As shown in Figure 4.29, the RAB is carried within the RRC connection between the RNC and UE over the radio interface and within the Radio Access Network Application Part (RANAP) protocol connection between RNC and CN over the Iu interface. In this association RNC acts as protocol converter between the RAN and CN. The RNC maps the requested RABs onto the RBs by using the current radio resource information and controls the lower layer services.

4.3.2.4 UTRAN Security

Radio connection security is another important function handled by the RNC. The RNC is involved in both integrity checking and ciphering mechanisms. The former is used for protecting the signalling connection between the UE and the UTRAN over the radio interface and the latter is used for protecting the user data transferred between the UE and the UTRAN.

The RNC ciphers the signalling and user data by using the defined integrity and ciphering algorithms. In this regard it needs to generate random numbers and maintain time-dependent counter values for the integrity checking of signalling messages. It also verifies and deciphers the received messages by using those algorithms. The algorithms and other UMTS security related issues are thoroughly addressed in Chapter 8.

4.3.2.5 UTRAN level Mobility Management

UTRAN level Mobility Management (MM) refers to those functions, which RNC handles in order to keep the UE in touch with the UTRAN radio cells, taking into account the user's mobility within UTRAN and the type of traffic or RAB(s) it is using.

As was mentioned in RRM subsection, the nature of traffic in the UMTS network environment is substantially different from the traditional circuit-switched type of traffic. This means that the UTRAN is able to share the radio resources for RABs, having different QoS. Therefore, more sophisticated mechanisms need to be utilised to exploit the radio resources efficiently and to meet the diverge QoS requirements of the RABs as well as is possible. To respond to this demand, the advanced RRM algorithms must be accompanied with more adaptive MM as it was for 2G.

As a result, the concepts of RRC state transition and hierarchical MM have been specified for UMTS, including UTRAN level MM, which is new compared to the GSM MM, which was taken care of by the CN subsystem. Based on this approach, during the radio connection the UE can have different states depending on the type of connection it may have to the UTRAN as well as the motion speed of the UE. It is the responsibility of the RNC to control the UE states by considering the UE mobility, the requested RAB(s) and its variation in terms of bit rate.

The cornerstones of the UTRAN level mobility are based on the concept of cell, UTRAN Registration Area (URA), Radio Network Temporary Identifier (U-RNTI) and the RRC state-transition model. In addition, the primary purpose of defining the different logical roles for RNC and specifying Iur interface was also to support the UTRAN internal MM.

A *URA* is defined as an area covered by a number of cells. The URA is only internally known in the UTRAN and hence it is not visible to the CN. This means that whenever a UE has a RRC connection its location is known by the UTRAN at the accuracy level of one URA. Every time the UE enters a new URA it has to perform a URA updating procedure. Having the interface between RNCs facilitates the URA area that can cover in principle different RNC areas.

Based on the 3GPP specification TS 25.401 *Radio Network Temporary Identities* (RNTI) are used as UE identifiers within UTRAN and in signalling messages between UE and UTRAN. There are four types of RNTI used to handle the UTRAN internal mobility, including:

- SRNC RNTI (S-RNTI)
- Drift RNC RNTI (D-RNTI)
- Cell RNTI (C-RNTI) and
- UTRAN RNTI (U-RNTI).

S-RNTI is allocated in association with RRC connection set-up by the SRNC to which the UE has the RRC connection. By this means the UE can identify itself to the SRNC and also the SRNC can reach the UE.

D-RNTI is allocated by a DRNC in association with context establishment and is used to handle the UE connection and context over the Iur interface.

C-RNTI, which is allocated when the UE accesses a new cell, it is a CRNC specific identifier and used to identify the CRNC to which the UE has the RRC connection. By this means the UE can identify itself to the CRNC and also the CRNC can reach the UE. This identifier is used when the UE is in Cell Forward Access Channel (FACH) state.

The U-RNTI is, on the other hand, allocated to a UE having a RRC connection and identifies the UE within UTRAN. U-RNTI is used as a UE identifier for the first cell access (at cell change) when a RRC connection exists for this UE and for UTRAN originated paging including associated response messages.

Allocating and handling of the above-mentioned identifiers is a primary responsibility of RNC.

Having defined the RNC roles, UE identifiers, and the concept of URA, then the question is how to combine this with the state transition in order to handle the UTRAN internal mobility and RRM. The main principles of the way of handling the RRC state transitions are shown in Figure 4.31 and can be summarised as follows:

- *No radio connection:* the UE location is known only by the CN (in accuracy of Location and/or Routing Area, which are CN level location area concepts and are described in Chapter 5). This means that the location information is stored in the network based on the latest MM activity the UE has performed with the CN.
- *Radio connection over common channels:* if the radio connection uses common channels, for example, FACH, and CPCH, the location of the UE is known in accuracy of a cell. The way in which this location information is updated is RRC procedure Cell Update. This state can be used when there is low bit rate data to be transferred between the UTRAN and the UE.
- *Radio connection over DCHs:* in this case the UTRAN has allocated dedicated resources for the connection (DPDCH) and Dedicated Physical Control Channel (DPCCH) as mini-

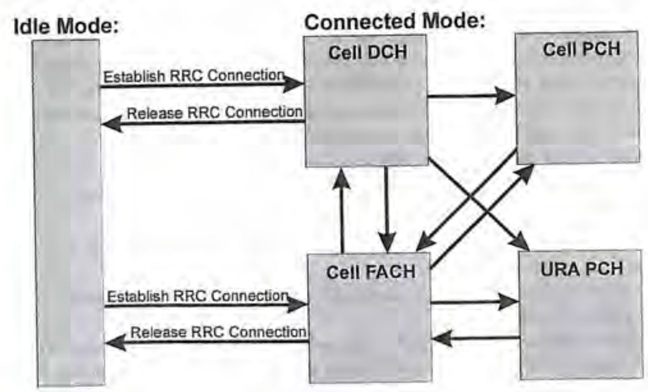


Figure 4.31. RRC state automaton

mum, number of DPDCH depends on the used bandwidth). The location of the UE is known on a cell level. Depending on the type of connection, different RRC procedures may occur. If the dedicated connection is used to carry a service using the highest QoS class (for example, circuit switched voice call is one of these) the UTRAN and the UE perform handovers. If the service the radio connection carries uses lower QoS class, for example, web surfing, allowing buffering and delays, handovers may not be completed as such. Instead, the UE uses Cell Update procedure to inform the UTRAN about its location, i.e. the place where the data should be delivered over the radio connection.

- *Radio connection in Cell PCH state:* Once the UE is in the Cell FACH or Cell DCH state but there is no data to be transferred, the UE state is changed to the Cell PCH state at which it can monitor the paging occasions based on the defined Discontinuous Reception (DRX) cycles and hence hear the paging channel. In this state the location of the UE is known at the accuracy of a single cell sometimes called the home cell.
- *Radio connection in URA PCH state;* Once the UE is in the Cell FACH or Cell DCH state and there is no considerable data to be transferred between the UE and the UTRAN, or the UE mobility is high, then its state can be transferred to the URA PCH state in order to avoid periodical cell update and to release the dedicated radio resources, respectively. In this case the location of the UE is known only at an URA level and hence in order to obtain the cell level location accuracy the UE should be paged by the UTRAN/RNC. This is the case, for example in association with mobile positioning process when cell based positioning method is used. In this state the UTRAN benefits from the advantage of having Iur interface between the RNCs and the URA concept.
- *Idle mode;*The RRC Idle Mode equals the state where the UE and the UTRAN do not have any radio connection, for example, the UE is switched off. In this RRC state the network does not have any kind of valid information concerning the location of the UE.

4.3.2.6 Database Handling

Like GSM BSC, the RNC contains stores of information that can be called a radio network database. This database is the data storage for cell information. The cell information for the cells RNC controls is stored in this database. The RNC then sends this cell-related information to the correct cell, which further distributes it by broadcasting over the Uu interface towards the UE. The radio network database contains plenty of cell-related information and the codes used in the cell are only part of this information.

The information in a radio network database can roughly be classified as:

- Cell Identification information: Codes, Cell ID number, Location Area ID and Routing Area ID of the cell.
- Power Control information: allowed power levels in the uplink and the downlink directions within the cell coverage area.
- Handover related information: connection quality and traffic related parameters triggering handover process for the UE.
- Environmental information: the neighbouring cell information (both GSM and WCDMA). These cell lists are delivered to the UE and the UE performs radio environment measurements as a preliminary works for the handovers.

4.3.2.7 UE Positioning

Another important function the RNC handles is to control the UE positioning mechanism in UTRAN. In this association it selects the appropriate positioning method and controls how the positioning method is carried out within UTRAN and in the UE. It also co-ordinates the UTRAN resources involved in the positioning of the UE.

In network-based positioning methods the RNC calculates the location estimate and indicates the achieved accuracy. It also controls a number of Location – Measurement Units/BSs (LMUs/BSs) for the purpose of obtaining radio measurements to position or help to position UE.

Since the UE positioning is seen as a value-added service in UMTS networks we discuss it more thoroughly in Chapter 7 in conjunction with the UMTS Services.

9

UMTS Protocols

In previous chapters the UMTS network has been discussed in the subnetwork and network element centric manner. The functional split between the network elements has been presented by allocation of major system-level management functions to the network elements. At the same time the interfaces between the different kind of network elements have already been introduced.

In this chapter we now take an interface-centric view of the UMTS network by focusing on the system protocols. The UMTS protocols are used to control the execution of network functions in a co-ordinated manner across the system interfaces.

9.1. Protocol Reference Architecture in 3GPP R99

Since the protocol specification work was divided between the various groups in the 3GPP technical organisation it was quite natural that each of the groups developed protocol reference architecture for those protocols for which it has got the mandate to do the specification work. This resulted in three major areas with protocol reference models of their own.

Before presenting a combined model for all UMTS network protocols, which will be followed throughout this chapter, let's have a look at each of those three reference models. It appears that among them some major protocol architectural concepts have been introduced and we can benefit from those when deriving the combined protocol architecture.

9.1.1. The Radio Interface Protocol Reference Model

The UTRAN radio interface is based on the WCDMA radio technology, which has some fundamental differences as compared to any 2G radio access technology. The key aspect to be controlled by the radio interface protocols is multiplexing of traffic flows of different kinds and different origins. To ensure effective control of multiplexing the layering of duties has been applied, thus resulting in the three-layer protocol reference model illustrated in Figure 9.1.

The layers are named simply according to their position in the architecture as radio interface layer 1, layer 2 and layer 3. Although the well-known OSI principle of layering has been applied in this design, these three layers cannot as such be considered as the three lowest layers of the OSI model. The layers have well-defined responsibilities and interfaces with each other, which makes it possible to name them as follows:

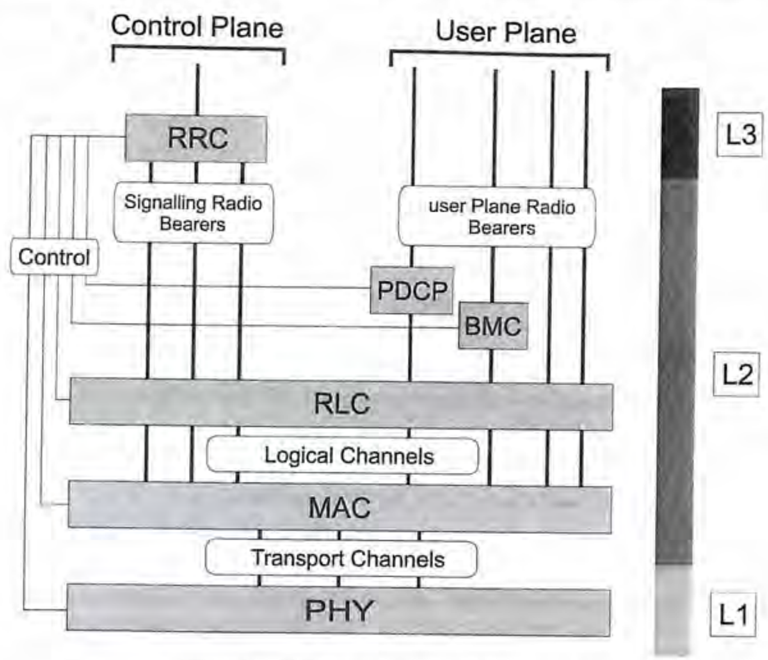


Figure 9.1. Radio interface protocol reference model

- L1 – radio physical layer
- L2 – radio link layer
- L3 – radio network layer

As shown in Figure 9.1 the physical layer provides its services as a set of WCDMA transport channels. This makes the physical layer responsible for the first multiplexing function: to map the flows from transport channels to WCDMA physical channels and vice versa. The mapping of transport channels onto physical channels was explained in Chapter 4.

The radio link layer is another multiplexing layer and it does make a major contribution to dynamic sharing of the capacity in the WCDMA radio interface. Instead of the wide variety of L1 transport channels this layer allows the upper layer to see only a set of radio bearers, along which different kinds of traffic can be easily transmitted over the radio. This UTRAN radio bearer is part of the UMTS system-wide bearer architecture, which was explained already in Chapter 1.

The Medium Access Control (MAC) sublayer controls the use of transport block capacity by ensuring that the capacity allocation decisions (done on the UTRAN side) are executed promptly on both sides of the radio interface. The Radio Link Control (RLC) sublayer then

adds regular link layer functions onto the logical channels provided by the MAC sublayer. Due to the characteristics of the radio transmission some special ingredients have been added to the RLC sublayer functionality.

For L3 control protocol (signalling) purposes the RLC service is adequate as such, but for domain-specific user data additional convergence protocols may be needed to accomplish the full radio bearer service. For CS domain data (e.g. transcoded speech) the convergence function is null, but for the PS domain an additional convergence sublayer is needed. This Packet Data Convergence Protocol (PDCP) sublayer makes the UMTS radio interface applicable to carry Internet Protocol (IP) data packets. Another convergence protocol (BMC) has been specified for message broadcast and multicast domains. The scheduling and delivery of cell broadcast messages to UEs is the main task of this protocol.

The separation of control signalling from user data is another key design criteria applied for a long time in protocol engineering. In this way the protocols used for control purposes become part of the system-wide control plane whereas protocols carrying the end-user data belong to the user plane.

The L3 control plane protocol the is Radio Resource Control (RRC) protocol. As shown in Figure 9.1 an RRC protocol entity on both the UE and UTRAN side has control interfaces with all other protocol entities. Whenever the protocol entity to be controlled by RRC is located in another UTRAN network element, there is a need to support this control mechanism with standardized protocols. In other cases the control interfaces are internal to a single UE or UTRAN element and hence not precisely standardised, but their existence is crucial to allow the RRC sublayer to carry out its task as executor of the radio resource management decisions.

9.1.2. UTRAN Protocol Reference Model

The access network, UTRAN, has the overall responsibility of WCDMA radio resources. UTRAN carries out the radio resource management and control among its own network elements and creates the radio access bearers to allow communication between UEs and the Core Network (CN) across the whole UTRAN infrastructure. This communication is structured according to the generic UTRAN protocol reference model illustrated in Figure 9.2. The model is generic in the sense, that the architectural elements – though not the specific protocols – are the same for all the UTRAN interfaces: Iu, Iub and Iur. The specific protocols are referred to in Figure 9.2 only by using their 3GPP specification numbers, which also indicates the idea of sharing the common protocol architecture.

The main elements of the UTRAN protocol architecture are layers and planes. The general layering principle is applied in the UTRAN design first of all to distinguish between two major parts of the protocol stack instead of going down into individual layers. Therefore, all the lower layer protocols together compose what is called transport network layer of UTRAN. The name transport is used here pretty much in its OSI sense, i.e. to cover the set of protocols allowing non-adjacent network nodes to communicate across an internetwork composed of possibly many different kind of subnetworks. Another, more pragmatic design aspect has been to include into the transport network layer all those protocols, which were selected from existing protocol suites instead of having to design them specially for UMTS purposes.

The other set of protocols on top of the transport network layer is – consistently with the radio interface protocol model – called radio network layer. These protocols have been

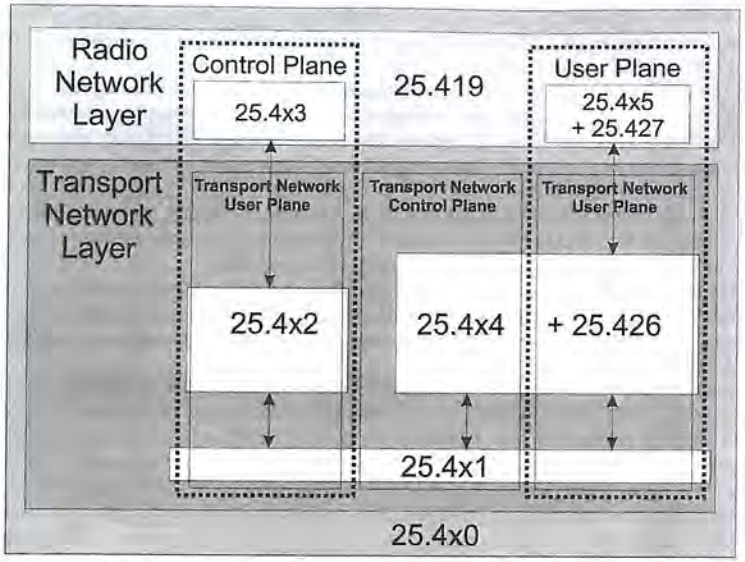


Figure 9.2. Generic UTRAN protocol reference model. Legend to the 3GPP specifications: $x = 1$ Iu protocol; $x = 2$ Iur protocols; $x = 3$ Iub protocols

carefully designed for the UMTS system and the common task for them is to control the management and use of radio access bearers across the various UTRAN interfaces.

Figure 9.2 also distinguishes between control plane and user plane protocols. This design aspect, which was already discussed before on radio interface protocols, has been applied over all UTRAN interfaces. Within the transport network layer the protocols have been selected keeping in mind the various properties required to support both control and user plane protocols in the most appropriate way. For control plane protocols this means reliability as a selection criteria and for user plane protocols, on the other hand, Quality of Service (QoS) support from the transport network has been targeted in protocol selection. The control protocols for the radio network layer has been specially designed to meet the radio and bearer control requirements. The control protocols for the UTRAN radio network layer are commonly named UTRAN Application Part (AP) protocols. On the other hand, the user plane protocols within the radio network layer, which share the common property of being responsible of efficient transfer of user data frames, are commonly knows as UTRAN frame protocols

The UTRAN AP and frame protocols together form the UMTS access stratum, which covers all those communication aspects, which are dependent on the selected radio access technology. Within the generic UMTS radio access bearers the non-access stratum protocols are then used for direct transfer of signalling and transparent flow of user data frames between UEs and the CN. This is achieved by the encapsulation of higher layer payload into UTRAN protocol messages.

The UTRAN control plane protocols have been designed to follow the client–server principle. Regarding the Iu interface UTRAN takes the role of a radio access server and the CN behaves as a client requesting access services from UTRAN. The same is true in the Iub interface, where BS is the server and its CRNC is a client. To some extent this applies also to the Iur interface, where actually the DRNC as a server provides the SRNC with a control service over the remote BSs. In the client–server based protocol design the behaviour of a server protocol entity is specified in terms of which actions it should take when receiving a service request from its client. On the other hand, under which conditions the client generates such service requests, may be left more unspecified.

The UTRAN protocol architecture is discussed in more details by Holma and Toskala (2001).

9.1.3. The CN Protocol Reference Model

The CN consists of the network elements, which provide support for the network features and end user services. The support provided includes functionality such as the management of user location information, control of network features and services and the transfer (switching and transmission) mechanisms for signalling and for user-generated information.

Within the CN the 3GPP R99 protocol architecture is derived from that of the GSM/GPRS system. Therefore five main protocol suites can be distinguished:

- Non-access stratum protocols between UEs and CN
- Network control signalling protocols between serving and home networks
- Packet data backbone network protocols
- Transit network control protocols
- Service control protocols

Due to the GSM/GPRS legacy these protocols have quite different specification backgrounds and origins.

First of all, the UMTS CN terminates the non-access stratum protocols from UEs. Among these the GSM/GPRS background is very evident. This is reflected by the 3GPP specification TS 24.008, which is an evolved version of its famous "ancestor" GSM 04.08. The design goal for the non-access stratum protocols has been to maintain compatibility between GSM/GPRS and UMTS systems in order to support combined networks (as shown in Figure 2.9) and dual-mode UEs.

The UMTS non-access stratum protocol stacks are shown in Figure 9.3. All of these protocols are carried over the signalling connection, which is established between UE and CN at the initial access and signalling connection establishment phase. In both PS and CS domains two sublayers can be distinguished in the non-access stratum protocols. The lower sublayer has to do with Mobility Management (MM); in the CS domain this protocol is called the MM protocol and in the PS domain it is called the GMM protocol to reflect the fact that it deals with the GPRS MM. On top of this common sublayer the more service specific Communication Management (CM) protocols are operated. The CM protocols and their control functions are as follows:

- Session Management (SM) protocol, which controls the establishment and release of packet data transfer sessions or Packet Data Protocol (PDP) contexts in the CN PS domain

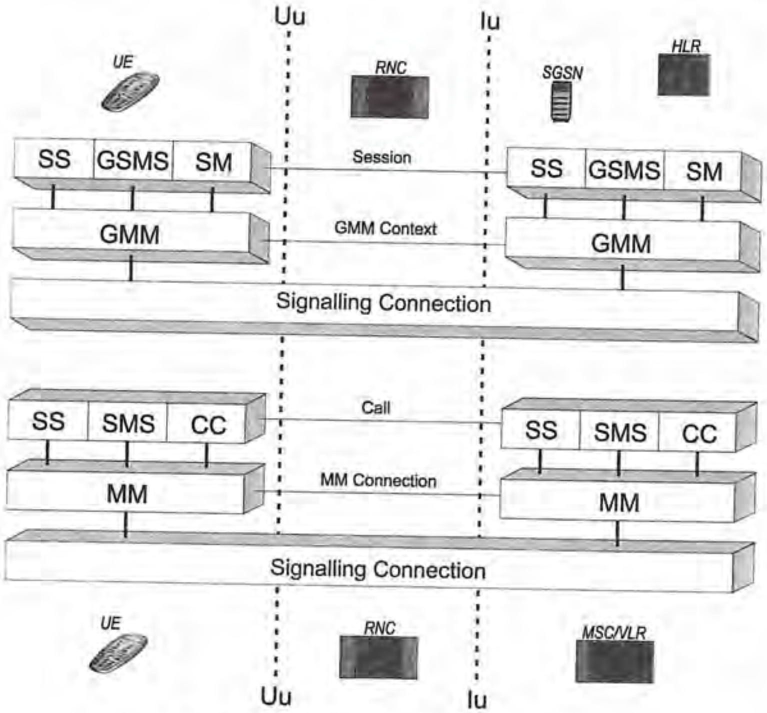


Figure 9.3. UMTS non-access stratum protocols

- Call Control (CC) protocol, which controls the establishment and release of circuit switched calls in the CN CS domain
- Supplementary Service (SS) protocol, which controls the activation and deactivation of various call-related and non-call related supplementary services
- Short message service (GSMS/SMS) protocol, which controls the delivery of short text messages to and from UEs.

The network control signalling between serving and home networks utilise the Mobile Application Part (MAP) protocol suite, which was originally designed for controlling the GSM circuit switched services. With the introduction of packet switched services by the GPRS subsystem it became necessary to extend the MAP protocol to the interfaces between GPRS support nodes and the home network nodes. From the 3G CN interfaces, which were illustrated in Figure 5.4, the following ones are controlled by the 3G version of MAP protocol:

- Interfaces C, D, E, F and G, which originate from the GSM system
- Interfaces Gc, Gr, Gf and Gs, which originate from the GPRS system

This means that normally both SGSN and GGSN must be equipped with the MAP protocol implementation for control plane purposes.

The MAP protocol follows a transaction-oriented communication scheme. Each transaction (e.g. registering the subscriber's location at HLR) is executed as a dialogue between CN nodes. This communication structure is created for the MAP sublayer by the Transaction Capabilities Application Part (TCAP) sublayer below. The TCAP further on utilises an SS7 based signalling transport network as the backbone for signalling network across network operator and national boundaries.

The protocol suite for the packet data transfer within the CN PS domain has also been adopted from the GPRS backbone network. It follows the internetwork protocol paradigm and IP protocol suite, which had fully established itself by the time GPRS specification was started. Actually only one GPRS-specific protocol was needed to be added. This GPRS Tunnelling Protocol (GTP) controls the communication across the PS domain backbone network in interfaces Gn and Gp (Figure 5.4).

The GTP protocol can also be modelled as two subprotocols: GTP-C for control signalling, which operates between GGSN and SGSNs and the user plane part GTP-U which extends from GGSN across the Iu interface to the UTRAN side. This termination of GTP-U communication on RNC (instead of SGSN) differs from the original GPRS specifications.

Any UMTS network has to interwork with external telecommunications and data communications networks. This interworking takes place across the transit network, which in the case of CS services is the international PSTN/ISDN backbone or in the case of PS services most often another IP backbone network. The protocols in the transit network boundary must therefore be aligned to those used in the transit network backbone. In the telephone network case the ISDN User Part (ISUP) protocol is the obvious choice and correspondingly IPv4 is still the dominating backbone protocol in IP data networks.

Within the UMTS CN value-added services should be available to subscribers not only within their home networks but also in visited serving networks. Therefore the service control protocol CAMEL Application Part (CAP) has been carried over from the GSM infrastructure as described in network evolution in Chapter 2. The CAP protocol is another transaction-based protocol and can therefore use the TCAP service and international SS7 signalling backbone just like the MAP protocol discussed above.

9.2. UMTS Protocol Interworking Architecture

The number and wide variety of UMTS system protocols described in Section 9.1 above may look confusing – and due to the legacy considerations which had to be taken into account in the protocol selection and design, it cannot be taken as a uniform and homogeneous protocol suite. In this section a combined protocol architecture model for the UMTS network is elaborated in order to guide the later examination of individual protocols in this chapter.

While traversing through the radio interface, UTRAN and CN protocols, we have already familiarised ourselves with two major protocol design aspects:

- Separation of (generic) transport aspects from (UMTS-specific) mobile networking aspects by layering.

- Separation of network control aspects from the user data aspects by planes.

Before starting to elaborate from this basis, let's take a closer look at two additional fundamental design considerations in building network-wide protocol architectures:

- Protocol interworking
- Protocol termination

Protocol interworking deals with protocols, which belong to the same layer but extend across multiple network elements and therefore comprise a set of protocols, which contribute to (distributed) execution of a common system-wide function. As examples of such functions one might consider radio access bearer set-up or location updating. The first requires interworking from UE up to the CN element and co-ordinated actions from the RRC protocol and the AP protocols within UTRAN. For location updating the "chain" of interworking protocols consists of the MM protocol (between UE and CN) and the MAP protocol between the CN nodes. Such a chain of distributed actions is called a (system) procedure among the protocol designers.

Protocol interworking is often designed as an extension to the above mentioned client–server model. Once an event is triggered at one edge of the network the protocol entity first discovering this external event becomes the client, which starts the procedure by requesting some action from its server. In a UMTS network the server seldom can satisfy such a request alone, but has to assume a role of another client and request other services from its server(s). This is how the procedure goes on, until those servers – often at the other edge of the network than where the triggering event took place – are reached, which cannot delegate the task any further but have to obey the commands themselves. In the case of radio access bearer set-up the procedure is initiated by a CN element and finally executed by BS and UE, which set up the radio link between them. In the location updating it is the UE, which initiates the procedure and finally the HLR has to record the new location into its database.

More complete examples of UMTS system level procedures are shown in Chapter 10, but this interworking of protocols within a layer is the key design aspect in setting up a layered view on the UMTS protocols in this Chapter, also.

Protocol termination can now be taken as the end point of a protocol chain, which interworks in the manner described above. Within a network element, where a protocol terminates, a system function or algorithm can be identified as the source/sink of the data or commands. For the MM protocols mentioned above in the location update example the termination points are therefore UE and HLR. Termination is sometimes discussed with respect to a single protocol, but then the termination points are simply the two end-points of the peer-to-peer protocol. Therefore, from the system-wide protocol architecture point of view, the termination of a set of interworking protocols, is much more significant.

Throughout the rest of this chapter the UMTS protocols will be described by dividing the UMTS protocol suite into three different layered subsystems. Each of them may be seen as a logical network of its own. This decomposition neither follows the functional network architecture described in Chapter 1 nor the strict structure of the 3GPP specifications, but rather takes a network-wide protocol-oriented view, where layering and end-to-end interworking are the main structuring principles.

First the UMTS protocol model is divided – in horizontal decomposition – into layers. The protocols within each of the layers operate across multiple interfaces and belong together in the sense of interworking described above.

Three layers can be distinguished as shown in Figure 9.4:

- Transport network layer
- Radio network layer
- System network layer

In this division the transport network (layer) is responsible for providing a general-purpose transport service for all UMTS network elements thus making them capable of communicating across all the interfaces described earlier. The UMTS system functionality is then distributed among the network elements by using the protocols within radio and system network layers, which are by definition UMTS system-specific. The radio network (layer) protocols ensure interworking between UE and CN on all radio access bearer related aspects. The system network (layer) protocols extend from UE until the transit network edge of the UMTS CN. They ensure interworking on UMTS communication service related aspects.

Within all three layers it is then possible to distinguish between control aspects and user data transfer aspects, which create the other structuring dimension to the UMTS protocol model shown in Figure 9.4. All protocols dealing with control aspects only belong to the control plane and those dealing with user data transfer belong to the user plane. The distinction between control and user planes is most visible within the radio and system networks, but within the transport network many protocols are common to the control and user plane. For optimisation reasons some of the transport network protocols are however selected specially to carry control or user data only.

The control plane protocols in different UMTS system interfaces interwork with each other to ensure system-wide control of communication resources and services. In a

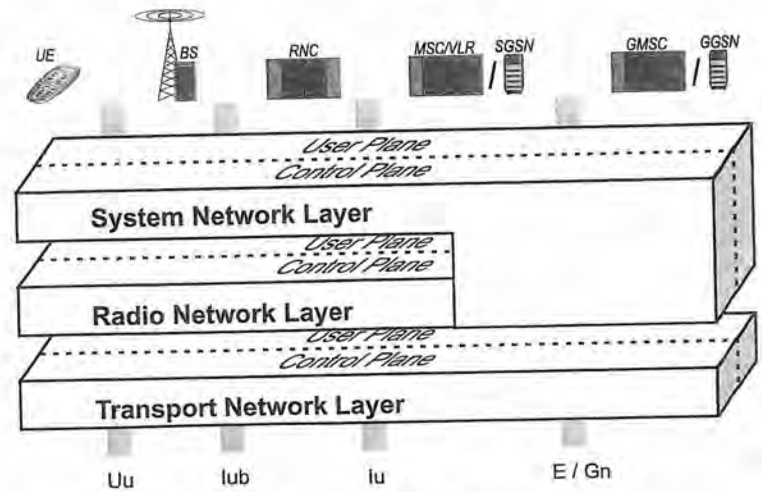


Figure 9.4. UMTS protocol internetworking architecture

similar way the user plane protocols interwork with each other to ensure the end-to-end flow of user data.

Based on the characteristics of the user traffic (within the CN) two different system domains are distinguished. In the Circuit Switched (CS) domain the user plane capacity is allocated in circuits for the whole duration of the service (e.g. a phone call). In the Packet Switched (PS) domain the user plane capacity is allocated in data packets. Such different requirements of these two domains are also reflected in the selection of transport network protocols and in the design of the system network protocols, which are therefore in this chapter often illustrated separately for both domains.

With this separation of functionalities the three different protocol subsystems can more easily be upgraded when new technologies become available. In fact such evolution is already taking shape in the introduction of alternative transport network protocols and the new IP Multimedia Subsystem (IMS) in release 4/5 of the 3GPP specifications.

9.3. Transport Network Protocol Aspects

Although ideally the communications capacity for user plane and control plane could be physically separate, it would lead to wasteful use of radio and terrestrial capacity in the network. Therefore the protocols providing general-purpose transport service for both user and control plane were designed as much as possible as a subsystem of its own, referred to as the transport network.

As the name suggests the UMTS transport network is actually a network within the UMTS network. A more detailed look at the transport network technology will further reveal that it is not just a single network, but actually consists of a number of networks, which when used in a UMTS system-specific way can been as a single (logical) UMTS transport network.

Close to the physical transmission of bits there is a transmission network, which except in the UMTS radio interface is typically based on digital transmission network technology. The (optical) trunk-line capacity of such a transmission network is shared by statistical multiplexing, which is controlled by the transmission network nodes. This transmission capacity is then utilised by the "second-degree" network on top of the transmission network. In the 3GPP R99 the preferred technology for this switching network is cell switching. This means that the data cells traverse through multiple switches before reaching their destination. Only at the edges of this switching network can we then find the actual UMTS network elements, which use this cell-switching technology to create the transport service between them.

The principle of composing an end-to-end transport network from a number of interconnected subnetworks is well known from the Internet protocol architecture. However, even if the UMTS transport network (in its 3GPP R99 level) is seen as an "internetwork", it does not follow such a uniform protocol structure as the IP internetworks. Instead the transport service within the different parts of the UMTS system – radio interface, RAN and CN – is designed and optimised for each part separately. As discussed below, there is a drive towards such a harmonised transport network architecture, which is based on IP networking with real-time transport capability.

Since the transport network itself provides switching and routing capabilities, it may also require an embedded control plane for the creation of transport network circuits or for distributing the routing information. These kinds of embedded control protocols are selected

according to the transport network technology itself and they are not discussed in any more detail here.

9.3.1. Transport Network Protocol Architecture

The transport network consists of the lowest layers of the UMTS protocol architecture thus providing facilities to transport and route both control and user traffic across all UMTS network interfaces. Figure 9.5 presents the most important transport network configurations in 3GPP R99, which are signalling transport network and packet transport network.

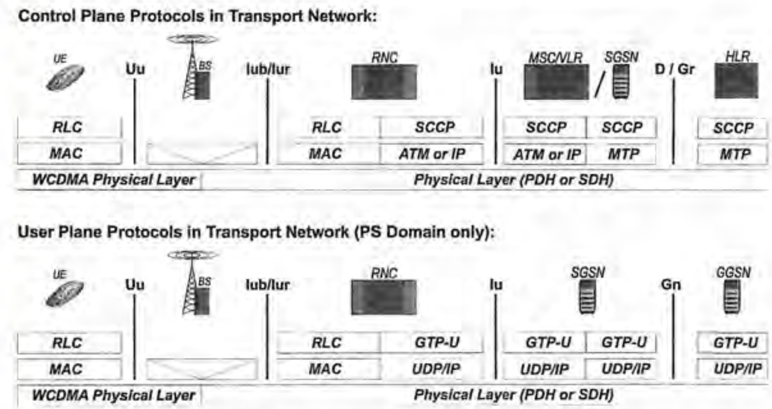


Figure 9.5. Transport network protocols in 3GPP R99

The physical layer controls the physical media through which both control signalling and user data traffic is transferred. The physical layer of the UMTS radio interface is based on the WCDMA radio technology as described in Chapter 3. The purpose of the physical layer protocol is to provide the upper layers with a set of WCDMA transport channels.

Over the terrestrial interfaces the physical layer may be formed by using general-purpose digital transmission technologies like Synchronous Digital Hierarchy (SDH) or Plesiochronous Digital Hierarchy (PDH). These are the most common alternatives available in transmission networks but in general any digital transmission system capable of providing stable bit flow with adequate bandwidth can be used. The SDH and PDH are the most obvious alternatives since they fulfil these basic requirements and are based on industrial standards.

The 3GPP release 1999 prefers ATM and its Adaptation Layers (AAL(n)) to be used especially in the UTRAN interfaces of the transport network protocol stack. The other recognised alternative, which is partly defined for release 1999 and is being standardised further in the release 4/5, is to use the IP suite over terrestrial interfaces.

The transport network protocol stack for the UTRAN radio interface as shown in Figure 9.5 is simplified a bit. Due to the UTRAN internal protocol structure and functional division the protocols MAC and RLC are terminated at the RNC. Therefore there is a need to carry the

MAC/RLC frames over the Iur/Iub interface. This part of the transport network is also based on ATM and its adaptation layers, but for simplicity reasons this is not shown in detail in Figure 9.5. Within UTRAN it is the SRNC, which is responsible for the radio interface related activities for a UE on the WCDMA transport channel level and the BSs actually only maintain the WCDMA physical channels.

9.3.2. Physical Layer Transmission Aspects

9.3.2.1. Physical Layer in the Uu Interface

The physical layer protocol controls the use of WCDMA physical channels (see Chapter 3) at the Uu interface (Figure 9.6). It provides the upper layer with a set of WCDMA transport channels, which characterise the different ways data can be transmitted over the WCDMA radio. The mapping between transport channels and physical channels is done by the physical layer protocol.

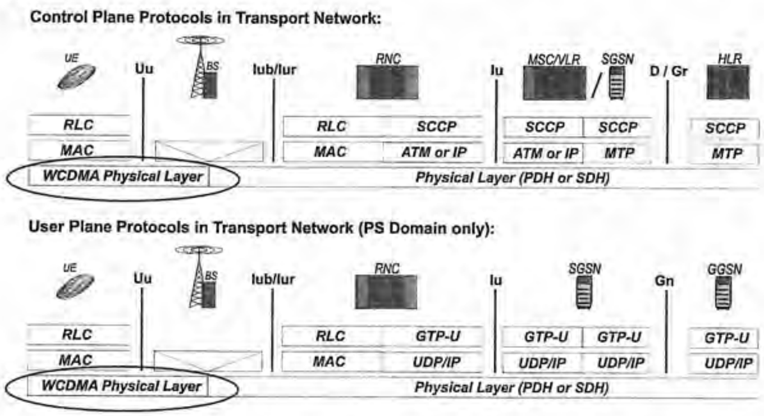


Figure 9.6. Physical layer in Uu interface

The physical layer provides bandwidth-on-demand service, which means that the transport channels support variable bit rates. As the lowest layer, the physical layer is responsible for multiplexing at the WCDMA frame level. To manage this kind of multiplexing the physical layer controls the transport formats used on transport channels. The *transport format* is a format for the delivery of a transport block set during each transmission time interval on a transport channel. *Transport block sets* are the physical layer SDUs composed by the MAC protocol entities, which use the physical layer service.

The *transmission time interval* determines how often the physical layer can accept data from the MAC layer. The transport format attributes are listed in Table 9.1.

The physical layer implements the multiplexing of transport blocks from different transport channels. In order to maintain the frame synchronization of the WCDMA radio transmission, the physical layer also performs segmenting of transport blocks into radio frames. Other functions of the physical layer are channel coding, interleaving and rate matching, which are all done before the radio frames are mapped for transmission over different physical channels.

Table 9.1 Transport format attributes for WCDMA-FDD

Dynamic attributes	Transport block size Transport block set size
Semi-static attributes	Transmission time interval Error protection scheme: type, coding rate, rate matching Size of CRC

As a comparison to the physical layer protocol in the GSM radio system it is worth noting that there is no ciphering at the WCDMA physical layer.

The physical layer protocol also attaches CRC checksum (0, 8, 12, 16, or 24 bits) into transport blocks to be transmitted over the radio interface. The receiving side the physical layer provides the transport blocks to the higher layer together with the possible error indication resulting from the CRC check.

Services provided by the WCDMA physical layer protocol are described in the 3GPP specification 25.302 and the detailed specification of the WCDMA physical layer protocol is covered by the 3GPP specifications TS 25.211–25.215. Detailed presentation of the WCDMA physical layer can be found in Holma and Toskala (2001).

9.3.2.2. The Physical Layer in Other Interfaces

In other, terrestrial, interfaces the physical layer can be implemented in many ways (Figure 9.7). The most obvious alternative is to utilise existing transmission, which typically consists

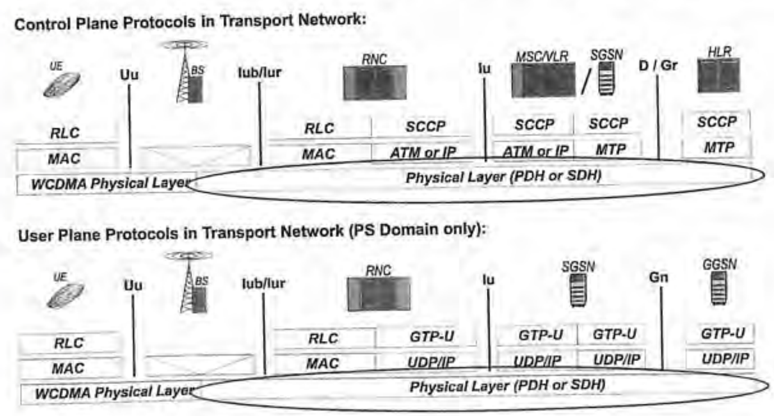


Figure 9.7. Physical layer in other interfaces

of E1/T1 trunk lines with Pulse Code Modulation (PCM) technology. This is especially the case when UMTS is implemented on top of existing GSM CN. On the other hand, UMTS requires the transmission to be flexible and able to provide adequate and changeable bandwidth in a physical sense without wasting transmission resources.

E1 trunk lines use timeslots, which in turn form a fixed timing structure and at the same time it fixes the transmission capacities offered for various transactions. Thus, transmission arrangement based on pure E1 trunk lines is not the best possible solution. E1 trunks can be used but without the timeslot structure. If used like this, E1 trunks offer bandwidth and with suitable multiplexing arrangements the adequate bandwidth is formed over several E1 connections. This arrangement is called inverse multiplexing. For example, the Iub interface between one BS and RNC could consist of four E1 trunks using inverse multiplexing. In this case, the Iub bandwidth available could be 4×2.048 Mb/s $= 8.192$ Mb/s.

Actually, Figure 9.8 represents the lowest hierarchy level of PDH. In PDH the multiplexing always occurs in groups of four, where the next higher hierarchy level is always a combination of four previous levels. Thus, the (European) PDH hierarchy levels are 2, 8, 34 and 140 Mb/s. Transmission signals having greater than 140 Mb/s bit rates are not specified in PDH. In addition to this, fibre optics interfaces are out of scope of PDH.

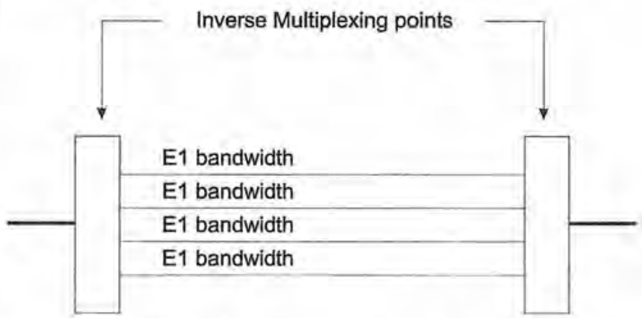


Figure 9.8. Inverse multiplexing – principle

PDH is very much based on 2.048 Mb/s (2 Mb/s) signals; the bit rate is always increased in steps of 2.048 Mb/s and also the system is synchronised on a 2.048 Mb/s level. The equipment performing higher level multiplexing use their own, plesiochronous synchronisation signals. This actually gives the PDH system its name. The weak point of PDH is this synchronisation scenario. Because the end-to-end synchronisation occurs on a 2.048 Mb/s level, it is very difficult to separate single 2.048 Mb/s or 64 kb/s or any other bandwidths from the higher-level bit stream.

SDH (Synchronous Digital Hierarchy), as a technology fixes the limitations of the PDH technology. To make it very simple, the SDH is a way to offer end-to-end synchronisation for both synchronous and plesiochronous transmission simultaneously. For instance, the PDH signal can be transported as SDH payload. In this case SDH forms a kind of transparent "tube" carrying the PDH signal from one end to another and the PDH signal is available for demultiplexing in the destination.

The smallest transmission unit in SDH is STM-1 (Synchronous Transfer Module # 1). STM-1 is actually a frame structure defining the way to synchronously transfer bits with a bit rate of 155.52 Mb/s. The STM-1 frame consists of header(s) and payload. The header part of the STM-1 frame holds synchronisation information and error correction information. The payload part of the STM-1 frame carries information to be transferred. The next hierarchy level in SDH is STM-4 containing $4 \times$ STM-1 which is equal to a bit rate of 622.08 Mb/s. The third SDH frame structure is STM-16 which is equivalent to a bit rate of 2488.32 Mb/s.

From the network point of view, the SDH is a more flexible technology since it offers the possibility to extract a single signal/bandwidth out from the main stream without difficult demultiplexing arrangements. This is true if the signals the SDH carries are synchronous ones. If the SDH carries plesiochronous signals they cannot be directly handled like the synchronous ones.

9.3.3. Traffic Flow Multiplexing Aspects

In order to use the transmission capacity provided by the physical layer in the most effective way multiplexing of traffic from different sources must be done within the transport network. Besides the WCDMA physical layer, which is itself capable of some multiplexing of traffic onto the WCDMA physical channels, multiplexing is a task given to the layers above the physical layer in the transport network.

In this section the protocol-oriented mechanisms for multiplexing are considered first at the UTRAN radio interface and then within the UTRAN and to some extent across the CN also.

9.3.3.1. Transport Network Protocols in the Uu Interface

The transport network protocols, which carry all control signalling and user data – for both packet and circuit switched traffic – over the Uu interface are

- MAC protocol, and
- RLC protocol

These two protocols work together as the radio link of the transport network.

Medium Access Control (MAC) Protocol

The MAC protocol is active at UE, BS and RNC entities (Figure 9.9). MAC is a layer 2 protocol and as part of the common transport network it provides services to both control and user plane. In the design of the MAC protocol both circuit switched and packet switched traffic, as well as signalling traffic considerations, have been taken into account from the very beginning. This is a major difference from the 2G packet radio protocols, which can be seen as packet-data specific add-ons on top of the normal circuit switched (time-slotted) 2G radio carriers.

MAC provides its service as a set of logical channels, which are characterized by what type of data is transported on them. There are four kinds of control channels and two kinds of traffic channels among those logical channels as described in Chapter 3.

MAC has the overall responsibility of controlling the communications over the WCDMA transport channels provided by the physical layer. In order to be able to share the capacity of the transport channels among the set of users, the MAC protocol uses transport blocks as units

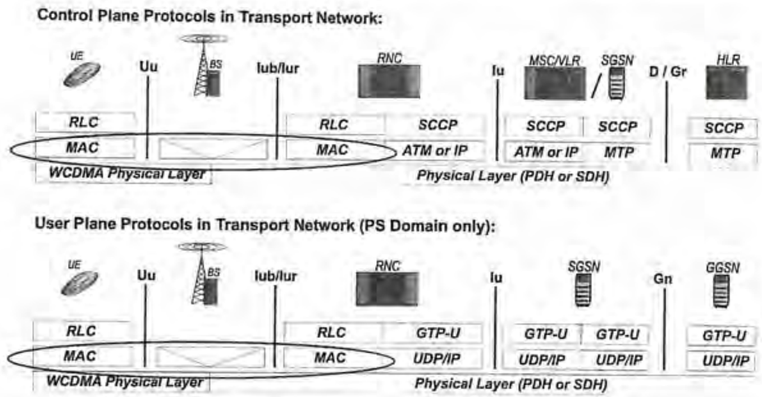


Figure 9.9. Transport network – medium access control

of transmission. The possible transport block sizes for each transport channel are offered to MAC by the WCDMA physical layer as a part of transport format set definition

The MAC protocol is capable of multiplexing PDUs from higher layers into transport block sets carried over common transport channels. Identification of UEs on common transport channels is performed by a temporary identity (allocated by RNC), which can be either Cell-Radio Network Temporary Identity (C-RNTI, 16 bits) or UTRAN Radio Network Temporary Identity (U-RNTI, 32 bits).

On dedicated transport channels MAC is capable of multiplexing by composing transport block sets from the higher layer PDUs. In this case all blocks in the set must belong to a single UE for which the transport channel has been dedicated and multiplexing is possible only if the QoS parameters are identical for the services supported by the dedicated channel.

This combination of different transport formats is based on the instantaneous bit rate of the traffic offered at any moment together with the power control considerations and results in transport format combinations as illustrated in Figure 9.10.

Due to the fact that MAC is also part of the user plane, it is a real time protocol. MAC has to meet the tight timing requirements of the physical layer. MAC must be ready to send a new set of transport blocks according to the transmission time intervals indicated by the physical layer. During each time interval MAC must select the optimal transport format combination based on characteristics of the data to be transmitted and on current traffic situation.

The basic service of the MAC layer is to provide data transfer of MAC SDUs between peer MAC entities. From the upper layers point of view it is worth noticing that this service is given without acknowledgements and segmentation.

Switching between common, dedicated and shared transport channels is also executed by the MAC layer, although the commands to switch the configuration are coming from the RRC layer.

MAC also collects statistical information about the traffic to be used by the RRC layer.

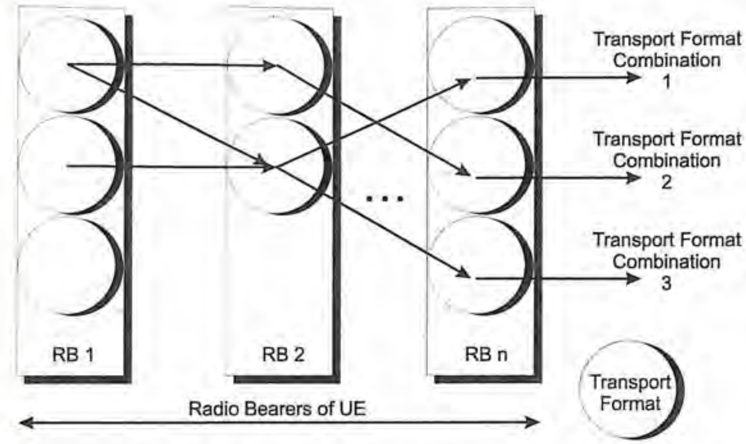


Figure 9.10. Transport format combinations in MAC protocol

and average. Also MAC status indication about underflow or overflow detected for each transport channel and measurement mode (event triggered, periodic) are reported to the RRC.

The MAC layer performs a ciphering algorithm of transparent mode RLC data. The MAC SDUs send on dedicated logical channels; DCCH and DTCH are ciphered with a UE specific key by using the block cipher algorithm KASUMI as described in Chapter 8.

The MAC protocol has only one PDU called data-PDU, which consists of a MAC SDU and MAC protocol header. The header is constructed according to the transport channel to be used. In some cases, e.g. when the UE has only one dedicated channel, the MAC header may be left out.

The new work items proposed within 3GPP for MAC extensions in release 4/5 include support for High Speed Packet Access (HSPA).

The complete specification of the MAC protocol is given in 3GPP specification TS 25.321 and a more detailed description of the MAC protocol can be found in Holma and Toskala (2001).

Radio Link Control (RLC) Protocol

The RLC protocol runs both in RNC and UE and it implements the regular (data) link layer functionality over the WCDMA radio interface (Figure 9.11). The RLC is active both in control and user plane simultaneously and it provides data link services for both circuit switched and packet switched connections.

The operational environment of the RLC depends on the plane discussed; in the case of the user plane the RLC is used by the PDCP protocol and in the case of the control plane the RLC is used by the RRC protocol. The lower layer protocol for the RLC is MAC, which provides data transfer services as MAC SDUs over logical radio channels. Because the MAC data transfer mode is unacknowledged, the RLC becomes responsible for the delivery of such higher layer

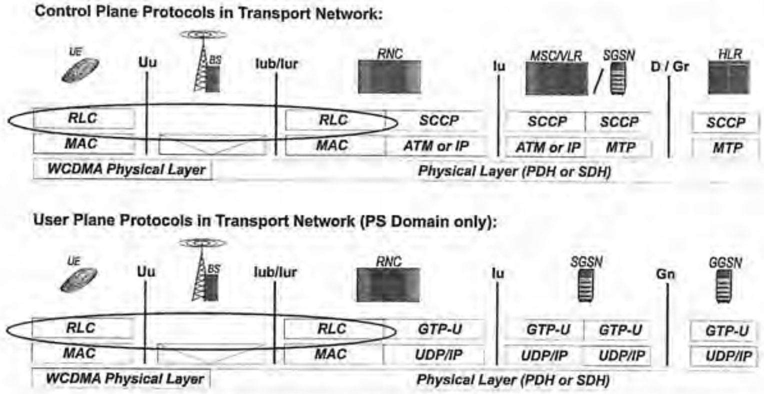


Figure 9.11. Transport network – radio link control

PDUs, for which reliability is needed. Also, because the MAC layer is not able to segment larger SDUs according to the transport formats available, RLC takes care of such functionality. Therefore the transport formats are visible at the RLC layer as shown in Figure 9.12.

RLC provides the data transfer service of higher layer PDUs as RLC SDUs. This service is called radio bearer service. Three modes of operation have been defined for RLC: transparent (Tr), unacknowledged (UM) and acknowledged (AM).

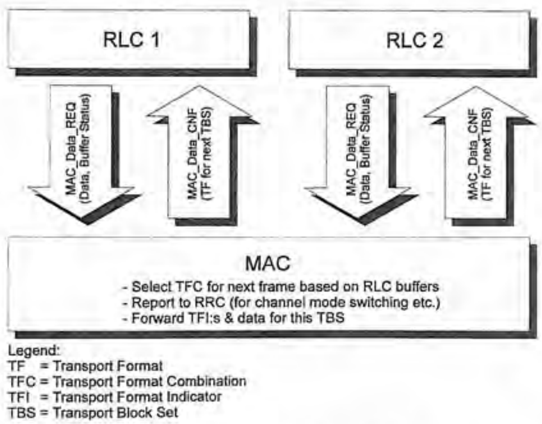


Figure 9.12. RLC interoperation with MAC on dedicated channels

In the *transparent mode* the RLC transmits SDUs without adding any protocol information. Segmentation and reassembly is still possible, but it has to be negotiated at the RRC layer during radio bearer set-up. This mode is used, e.g. for streaming class services.

In the *unacknowledged mode* the RLC transmits SDUs without guaranteeing delivery to the peer entity. This mode is used by some RRC control procedures, where acknowledgement is taken care of by the RRC protocol itself.

In the *acknowledged mode* the RLC transmits SDUs and guarantees the delivery to its peer entity. The guaranteed delivery is ensured by means of retransmission. In case RLC is unable to deliver the data correctly, the user of RLC at the transmitting side is notified. This mode is used for packet switched data transfer over dedicated logical channels.

The RLC has numerous functions to perform on its layer such as:

- Segmentation and reassembly
- Concatenation
- Padding
- Error correction
- In-sequence delivery of SDUs
- Duplicate detection
- Flow control
- Sequence number check
- Protocol error detection and recovery
- Suspend/resume functionality
- SDU discard
- Ciphering

Segmentation and reassembly means the adjusting of variable-length higher layer PDUs into/from RLC PDUs. The RLC PDU size is adjustable to the actual set of transport formats.

Concatenation is used when the contents of an RLC SDU do not fill an integer number of RLC PDUs. Then the first segment of the next RLC SDU may be put into the RLC PDU in concatenation with the last segment of the previous RLC SDU. When concatenation is not applied and the remaining data to be transmitted does not fill an entire RLC PDU of a given size, the remainder of the data field shall be filled with padding bits. Padding can be replaced by piggybacked status information for the reverse link.

Many functions of RLC have to do with error detection and correction. The bit errors may actually be detected by the physical layer CRC check, but the RLC layer is responsible of error recovery. The most effective error recovery is provided by retransmission in acknowledged data transfer mode. The RLC protocol can be configured by RRC to obey different retransmission schemes, e.g. Selective Repeat, Go Back N or Stop-and-Wait.

In-sequence delivery of SDUs preserves the order of higher layer PDUs that were submitted by using the acknowledged data transfer service. If this function is not used, out-of-sequence delivery may happen.

The RLC is able to detect duplicates of the received RLC PDUs and can ensure that the resultant higher layer PDU is delivered only once to the upper layer. Flow control allows an RLC receiver to control the rate at which the peer RLC entity may send information.

Sequence number check function may be used in unacknowledged mode to guarantee the integrity of reassembled PDUs. It provides a mechanism for the detection of corrupted RLC

SDUs through checking the sequence number in RLC PDUs when they are reassembled into a RLC SDU. A corrupted RLC SDU will be discarded.

Protocol error detection and recovery functionality detects and recovers from errors in the operation of the RLC protocol. The RLC reset procedure is used to recover from an error situation. In case of an unrecoverable error the RLC entity notifies the RRC layer.

The RLC may suspend and resume the data transfer under request from the RRC.

The SDU discard function allows the discarding of the remaining RLC PDUs of that SDU from the buffer on the transmitter side, when the transmission of an RLC PDU does not succeed for a long time. The SDU discard function allows the avoiding of buffer overflow. There are several alternative operation modes of the RLC SDU discard function.

Ciphering is performed in the RLC layer for those radio bearers, which use unacknowledged or acknowledged mode. Ciphering is done with a UE specific key by using the block cipher algorithm KASUMI as described in Chapter 8.

The complete specification of the RLC protocol is given in 3GPP specification TS 25.322 and a more detailed description of the RLC protocol can be found in Holma and Toskala (2001).

9.3.3.2. Transport Network Protocols in Other Interfaces

Unlike in the radio interface the transport network protocols for the UMTS terrestrial interfaces have not been designed specifically for UMTS transport purposes, but instead the 3GPP standardisation bodies have selected them among the existing protocol suites. As in the case of all transport network protocols the focus in this selection has been on the ability to multiplex traffic from different end users and – to be more precise – from different UMTS bearers taking into account their QoS characteristics.

This – together with some standardisation politics, which inevitably is also present – has led to the selection of two major protocol suites: one coming from the broadband telecommunications, the ATM protocol family, and the other coming from the Internet data communication networks and known as the IP protocol family. In the 3GPP release 1999 the ATM is the dominating transport network technology on the UTRAN side and the same is true for IP-protocol technology in the PS domain on the CN side. Within the CN CS domain the transport network will continue to be based on the time-slotted PCM trunking network.

In order to make the ATM and IP protocols suitable for the UMTS transport network service, both of them have required some adaptations as compared to their use in their native networking environments. In the case of ATM such adaptations mostly have to do with conversational speech transmission capability whereas the IP protocol stack has been extended to meet the multiplexing of user data and signalling transport requirements. The single most significant extension to the IP transport is adoption of the GPRS tunnelling protocol to carry user data packets between UTRAN and CN nodes.

ATM Transport

The basic idea in ATM (Figure 9.13) is to split the information flow to be transferred into small pieces (packets), attach address tags to those packets and then transfer the packets through the physical transmission path. The receiving end collects received packets and forms original-like information flow from the contents of the packets. The packet containing transmitted information is called the *ATM cell*.

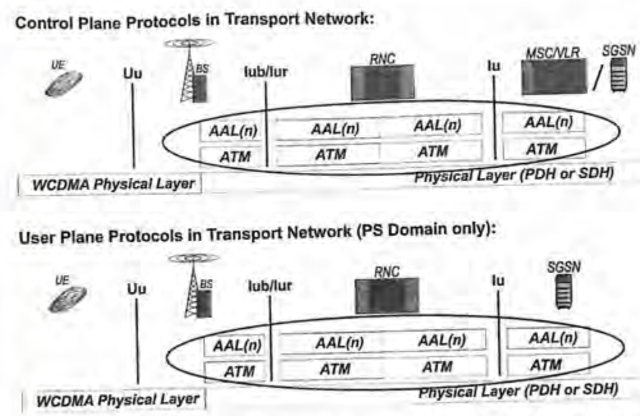


Figure 9.13. Transport network – asynchronous transfer mode

An ATM cell as shown in Figure 9.14 consists of two parts, a 5-byte-long header (address information) and payload (transmitted information). When comparing to "conventional" protocols and messages, the header is very short. This sets some limitations on what can be done but on the other hand, the information transfer effectiveness is high: the addressing overhead is $5/(5 + 48) \approx 9.5\%$. Another aim has been to establish a very lightweight transmission system without any extra "bureaucracy". Because of this, the payload of an ATM cell is *not* protected with check sum method(s). Nowadays this is possible because the transmission networks carrying ATM traffic are of high quality and the terminals used are able to perform error protection themselves if required.

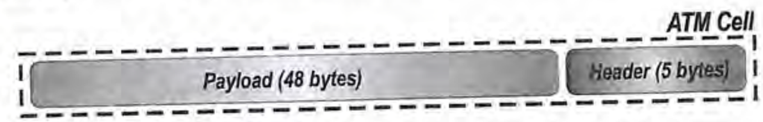


Figure 9.14. ATM cell structure

The header of an ATM cell contains some address information as illustrated in Figure 9.15. The most essential items are:

- VPI (Virtual Path Identifier): the identifier for a Virtual Path (VP), or more generally, an identifier for a constantly allocated semi-permanent connection.
- VCI (Virtual Channel Identifier): an identifier for a Virtual Circuit (VC). This field is long because there may be thousands of channels to be identified within one VP. For instance, multimedia applications may require several VCIs simultaneously, one VC per each multimedia component.

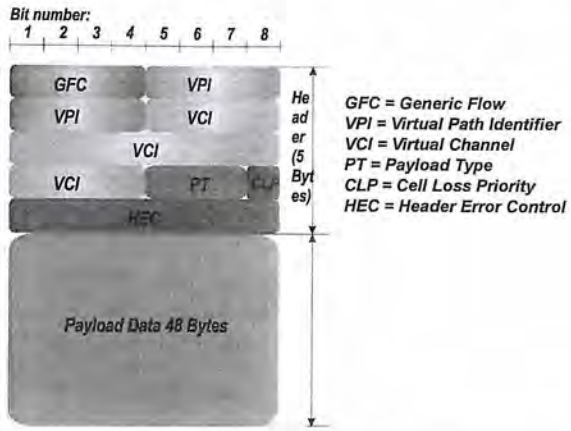


Figure 9.15. ATM cell header structure

- PT (Payload Type): this indicates whether the 48-byte payload field carries user data or control data.
- CLP (Cell Loss Priority): this is a flag indicating if this ATM cell is "important" or "less important". If CLP = 1 (low priority/less important) the system may lose this ATM cell if it has to.
- HEC (Header Error Control): in ATM, the ATM cell header is error protected.
- Reason: a failure in the ATM cell header is more serious than in payload because due to header error the ATM cell may be delivered to the wrong address, for instance. The error correction mechanism used is able to detect all errors in the header and one failure can be corrected.

Figure 9.16 illustrates the transmission path of ATM. One ATM transmission path may consist of several virtual paths, which further on contain virtual channels.

A virtual path is a semi-permanent connection simultaneously handling many virtual connections/channels. Actual data is transferred in ATM cells over the virtual channels.

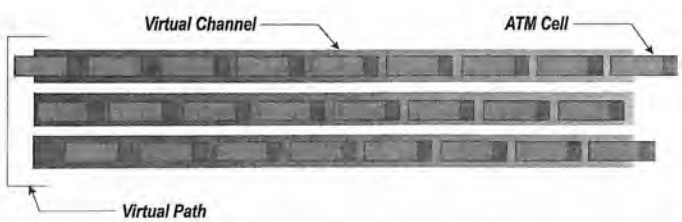


Figure 9.16. Virtual Path (VP) and Virtual Channel (VC)

From the point of view of the UMTS system, an ATM transmission path is, say, between the BSs and the RNC. If a loop transmission is in question, the transmission path contains many virtual paths (one per BS) and the virtual channels in the virtual path are set up on a per call basis. The bandwidth of the virtual channel varies depending on the bearer service used.

The ATM layer as such consists of fairly simple transport media and in theory suitable for transmission purposes as such. In practise, the ATM layer must be adapted to the higher protocol layers and the lower physical layer. ITU-T has defined what are called ATM service classes with AALs.. The original idea was that each service class from A to D should correspond to one AAL from 1 to 4. As time went by, the original idea disappeared as can be seen from Figure 9.17.

The service classes of the ATM are

- Constant Bit Rate Service (CBR)
- Unspecified Bit Rate Service (UBR)
- Available Bit Rate Service (ABR)
- Variable Bit Rate Service (VBR)

Any transparent data transfer may use the CBR and the resources are allocated on a peak data rate basis. The UBR uses free bandwidth when available. If there are no resources available, queuing may occur. The ABR is used when the user service has a minimum bit rate defined. Otherwise the bandwidth is used as in UBR. The VBR provides a variable bit rate based on statistical traffic management.

The adaptation layers for ATM are as follows:

- AAL1 offers synchronous mode, connection oriented connection and constant bit rate for the services requiring this kind of adaptation.

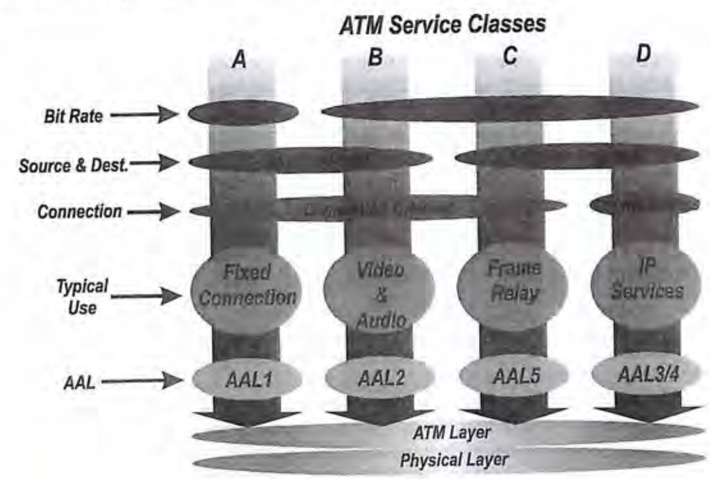


Figure 9.17. ATM Adaptation Layers (AAL)

- AAL2 offers synchronous mode, connection oriented connection with variable bit rate for the service using this adaptation.
- AAL3/4 offers connectionless and asynchronous connection with variable bit rate.
- AAL5 offers asynchronous mode, connection oriented connection with variable bit rate.

As shown in Figure 9.18, AAL is divided into two sublayers, Convergence Sublayer (CS) and Segmentation And Re-assembly (SAR) sublayer. The CS sublayer adapts AAL to the upper protocol layers and the SAR splits data to be transmitted into suitable payload pieces and in receiving direction it collects payload pieces and assembles them back to original dataflow. Depending on the case, the CS sublayer may be divided further into smaller entities.

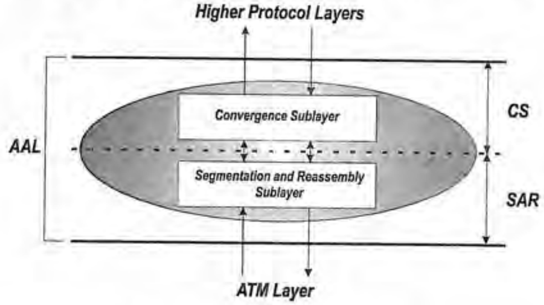


Figure 9.18. General structure of AAL

From the point of view of the UMTS transport network, AAL2 and AAL5 are the most interesting alternatives. AAL2 is seen as a suitable option for circuit switched user plane connections and AAL5 is seen as suitable for control protocol exchange. A more detailed list of these AAL variants in different UTRAN protocol interfaces is given in Table 9.2.

Table 9.2 ATM adaptation layers used at the UTRAN interfaces

AAL5	Iu: CS C-plane, PS C/U-plane, internal C-plane of the transport network Iur: C-plane, internal C-plane of the transport network Iub: C-plane, internal C-plane of the transport network
AAL2	Iu: CS U-plane Iur: U-plane Iub: U-plane

Within UTRAN the ATM adaptation layer AAL5 is used to carry all control protocols as well as the PS domain user data at the Iu interface. On the other hand AAL2 is used as the common carrier for user data in all interfaces, except the PS domain user data at the Iu interface.

The sublayers of the AAL protocol stack in different UTRAN interfaces are not described

in detail here, but can be found in the 3GPP specifications numbered in Figure 9.2. The key observation in the selection of these convergence protocols is that in all control plane protocols for the PS domain the convergence may be accomplished in two alternative ways. Either the convergence protocols specified by ITU-T for signalling transport over AAL5 are selected or IP-over-AAL5 is used. Since also the U-plane convergence for the PS domain at the Iu interface is based on User Datagram Protocol (UDP) over IP, this means that for the PS domain the transport network can be completely based on an "IP-over-ATM" implementation option. This serves as a migration path towards wider utilisation of IP-based transport in the 3GPP release 4/5.

IP Transport for User Data (GTP-U)

Within the 3GPP R99 set of specifications IP-based transport is widely applied only within the CN PS domain backbone network and at the Iu interface for PS domain user plane traffic as shown in Figure 9.19. The major protocol providing the transport network service in these interfaces is the GPRS Tunnelling Protocol for packet data User Plane (GTP-U). As the name suggests this protocol has been adopted from the 2G GSM/GPRS system, but with a significantly different architectural design choice. In 2G the GTP-U protocol is used only between the 2G-GSN network elements, but in 3G is was extended to reach RNC across the Iu–PS interface also.

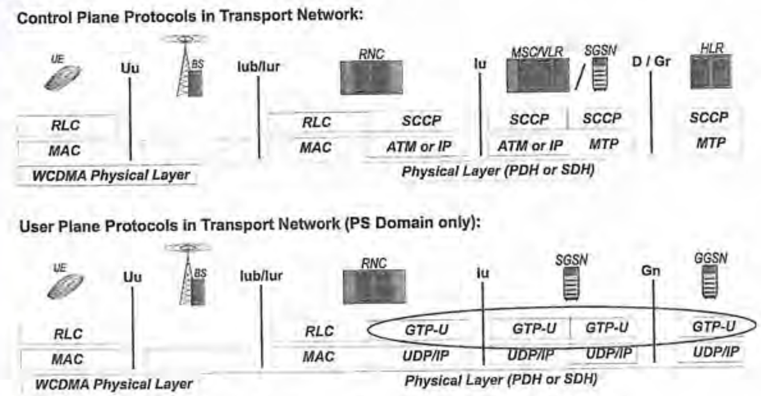


Figure 9.19. GPRS tunnelling protocol for user plane

GTP-U is a transport network protocol supporting user plane data transfer in the UMTS network. GTP-U belongs to the packet switched domain in the UMTS network and is located in RNC on the UTRAN side, as well as in SGSN and GGSN on the CN side.

GTP-U operates on three different interfaces; Gn, Gp and Iu-PS. The Gn interface is between GSNs (i.e. SGSNs and GGSNs) belonging to the same UMTS packet domain network. The Gp interface is used in network interworking between two SGSNs or between SGSN and GGSN belonging to different UMTS packet domain networks. Iu-PS is the interface between RNC and SGSN.

On all interfaces GTP-U operates on top of the UDP/IP protocol family. User Datagram Protocol (UDP) provides connectionless message transfer between two IP network nodes. UDP addressing is used to identify GTP-U end points within source and destination network elements. UDP also provides checksum mechanism to detect transmission errors inside data packets. As it is generally known the main service of the IP is the routing of messages between source and destination network elements. This routing is based on IP network addresses (both IPv4 and IPv6 may be used within the UMTS transport network).

GTP-U provides connectionless data transfer services for upper layers. GTP-U allows multi-protocol user data packets to be tunnelled across the Iu-PS, Gn and Gp interfaces. Because of the encapsulation and tunnelling mechanisms it is possible to transfer user data packets utilising different routing protocols in the packet domain IP backbone network. Also UMTS network elements and protocols transferring user data packets between UTRAN and GGSN need not be aware of different addressing mechanisms at the IP layer in order to be able to associate user data packets to specific PDP contexts.

Since the main purpose of the GTP-U protocol is to transfer user data, it has been optimised to that specific task. GTP-U multiplexes packets received from one interface (e.g. Gi at the GGSN) and addressed to several destinations (different UEs). GTP-U receives the user data packet from the external packet network, interprets the destination of the packet and passes the packet onto the next node along the path. Tasks for setting up the tunnels between GTP-U endpoints have been excluded from GTP-U and control plane protocols are used for those purposes. On the Gn interface GTP-C (GTP for control plane) and on the Iu-PS interface RANAP protocols are used to control the set-up of GTP-U tunnels.

GTP-U protocol entities have to interwork with other protocol entities along the way the data packets are transported. In GGSN the GTP-U protocol entity communicates with the packet relay functionality located at the edge of the UMTS network, i.e. the Gi interface. At the Iu-PS side of an SGSN the GTP-U protocol entity interworks with another GTP-U entity at the Gn side. Within RNC the GTP-U protocol entity terminates the tunnel and forwards the packets to the frame protocol entity at the radio network layer.

The interworking between GTP-U and the neighbouring protocol entities is quite simple. After all, all parameters related to user data packet transfer have been negotiated in advance by the control plane. User data packets are just passed to the next entity, which is then able to handle the packet according to its predefined PDP context parameters.

The main functions of the GTP-U protocol are

- Data packet transfer
- Encapsulation and tunnelling
- Data packet sequencing
- Path alive check

The maximum size of the user data packet is 1500 octets. User data packets containing 1500 octets or less shall be transmitted as one packet. In case the user data packet received from the external network by GGSN exceeds the 1500 octet limit, GGSN shall fragment or discard the user data packet depending on the PDP type. Since IP fragmentation is inefficient and does not tolerate transmission errors, fragmentation should be avoided. Thus all links between network elements containing GTP-U should have Maximum Transmission Unit (MTU) values exceeding the 1500 octets plus size of GTP-U, UDP and IP headers.

Each user data packet is encapsulated before being transmitted into the UMTS packet domain network. The encapsulation adds a GTP protocol header containing tunnelling information into every user data packet. Tunnelling information includes a 32-bit Tunnel Endpoint Identifier (TEID), which serves two important purposes. Firstly, TEID is used to address a PDP context inside a tunnel endpoint. Hence, it also indirectly refers to different UEs, which have one or more PDP contexts active in the UMTS network. Secondly, tunnelling enables multiplexing of user data packets destined for different addresses into a single path that is identified by two IP addresses. This removes the need to understand different routing protocols the user may use in the UMTS network and thus makes the UMTS network simpler. GTP-U tunnelling makes it also simpler to allow end users to utilise new routing protocols.

As an optional feature GTP-U may preserve the order of user data packets between RNC and GGSN. During PDP context establishment at the control plane the reordering may be negotiated. In case user data packet reordering is applied, the GTP header shall contain a 16-bit sequence number, which GTP-U uses to determine if user data packets are received in the correct order or whether it has to wait for the missing packets. However, in case the missing user data packet is lost, the data packet sequencing mechanism does not provide a mechanism to recover a lost packet. It is up to the other layers to recover from the lost user data packet.

GTP may check at regular time intervals if the peer GTP is alive by sending an echo request message to the peer GTP. The peer GTP shall respond to an echo request message with an echo response message. On the Gn interface one may think that the path alive check procedure belongs to GTP-C but on the Iu-PS interface GTP-U is the only GTP element and thus handles the path alive check procedure. On the Gp interface the path alive check is similar to the Gn interface.

The main procedures of the UMTS network where the GTP-U protocol is used are first of all the user data packet transfer between UTRAN and SGSN as well as between SGSN and GGSN. When a routing area update procedure requires to switch a GTP-U tunnel from one SGSN to another, the GTP-U is used to relay user data packets not yet sent to the UE from the old SGSN to the new SGSN. GTP-U is also used as part of the SRNS relocation procedure to tunnel user data packets not yet sent to the UE from the source RNC to the target RNC via the pair of SGSNs attached to the RNCs.

The GTP-U protocol is defined in the 3GPP specification TS 29.060.

9.3.4. Signalling Transport Within the CN

Due to the inherited background, i.e. GSM and its elements and functionalities, the 3GPP R99 implementation still includes a protocol stack based on the Common Channel Signalling System 7 (SS7). SS7 is used for signalling transport within the CS domain of the CN as illustrated in Figure 9.20.

The SS7 protocol stack is divided into a physical layer, signalling link layer (MTP2) and signalling network layer (MTP3 and SCCP). The SS7 physical layer is based on PCM transmission, which makes it possible to multiplex user data (circuit switched voice) and signalling in a time-slotted manner. One of the 64 kb/s time-slots on a PCM trunk line is typically reserved for the signalling traffic. Besides this structure originating from the narrowband SS7 a broadband version of the protocol stack is also used within UTRAN. In this case MTP3 is run on top of the ATM protocol stack, which leads to a new adaptation of this protocol called MTP3b.

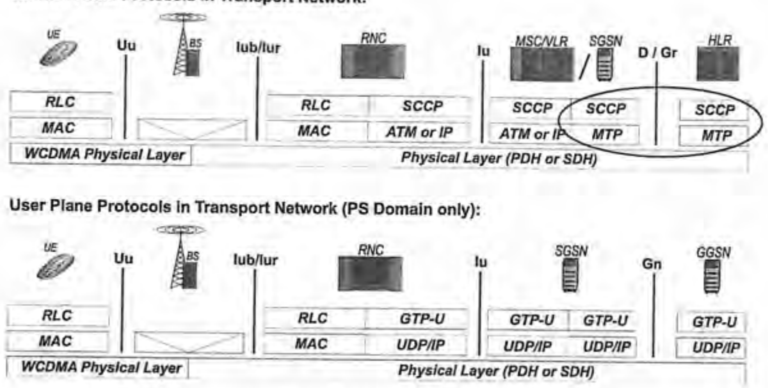


Figure 9.20. SS7 Transport in CN

The SS7 basic protocol stack provides signalling connections, their control and basic signalling routing functionality. For various application-oriented signalling needs one must add "intelligence" by adding more protocols to the protocol stack. For example, to offer connection-oriented and connectionless services within an SS7 environment, an additional protocol called Signalling Connection Control Part (SCCP) is required and this protocol sits on top of the basic SS7 protocol stack.

The SS7 uses three kinds of messages: FISU (Fill-In Signalling Unit), LSSU (Line Status Signalling Unit) and MSU (Message Signalling Unit). The signalling channel must be populated all the time, hence, if there isn't any information to send, the signalling node sends FISU. FISU does not contain any upper layer information, it only contains sequence number and indicator bits for acknowledgement purposes. LSSU is sent when the SS7 nodes need to negotiate/change a signalling channel status or they have to inform each other about other maintenance activities. MSU is sent when there is some upper layer information to be delivered.

As illustrated in Figure 9.21 the SS7 message is always started with a Frame Mark (*F*). *F* is a fixed bit pattern 01111110 and after this the receiving signal node is expecting some SS7 information to be received. After *F* the SS7 message contains sequence numbers in both back and forward directions (BSN and FSN) and indicator bits for the same directions (BIBand FIB). With these four fields the receiving node is able to perform message acknowledgement activities. In SS7 any message can acknowledge another; for instance a FISU can acknowledge an MSU.

The Length Indicator (LI) field indicates how many octets long the SS7 message is. Up to this point, all the fields in the message are related to data link layer activities.

The next part, Service Information Octet (SIO), indicates the user (protocol) to which this message is addressed. SIO may contain a bit patter indicating that, for instance, ISUP is the protocol handling this MSU. The real, higher-level message is in the Service Information Field (SIF). At the beginning of this field the MSU contains addressing information for the

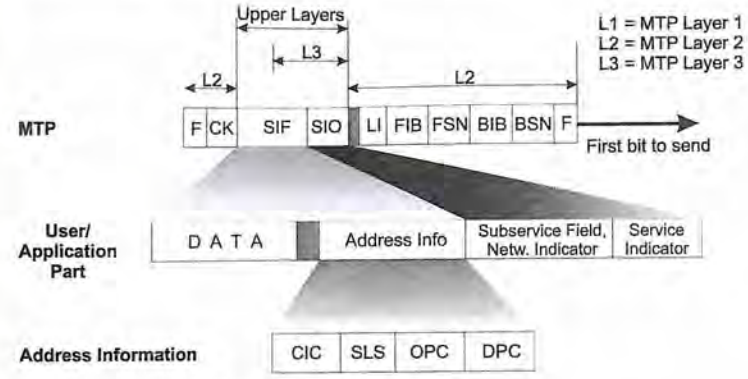


Figure 9.21. SS7 Message Structure (MSU)

MTP layer, which is responsible for message routing. In the SS7 network addresses of the signalling points are given as Signalling Point Codes (SPC). The MTP routing facility checks the Originating Signalling Point Code (OPC) and Destination Signalling Point Code (DPC) and if the DPC is the same as that defined for the current MTP node, the message is interpreted to be terminated in this signalling point. If different, the message is re-routed towards the correct DPC.

In addition, the SIF address part contains identification for used signalling channels, Signalling Link Selection (SLS) and circuit concerned, Circuit Identification Code (CIC).

The basic element of SS7 signalling information transfer is signalling link. Signalling link is a data link layer connection between two signalling nodes as shown in Figure 9.22. Both nodes identify a signalling link with a unique number, Signalling Link Code (SLC). The SLC (or SLS) should be same in both ends of the signalling link and checking for this is one of the items the LSSU messages are used for.

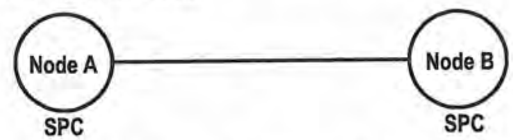


Figure 9.22. SS7 signalling link

One signalling link between two nodes is able to handle a certain amount of signalling traffic but sooner or later more links will be required. The set of signalling links between two signalling nodes, as shown in Figure 9.23, is called Signalling Link Set (SLS). For proper signalling link selection, every signalling link within a SLS must have unique SLC. The signalling traffic is carried through all the signalling links within the SLS.

Normal practise is that a signalling session (for instance signalling related to ISUP call set-up) is carried through by using the same signalling link for all messages. If *load sharing* is

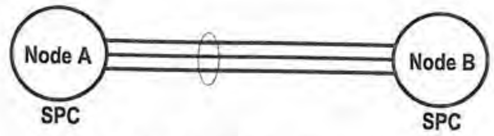


Figure 9.23. SS7 signalling link set

used, the MTP level is able to distribute messages from one signalling session over several links.

The SS7 makes it possible to have a situation where the actual traffic path is geographically different from that of the related signalling. This is, two nodes may have direct traffic connections, but the signalling related to those connections is handled through other nodes.

The signalling node taking care of re-routing of the messages in this case is called STP (Signalling Transfer Point), please refer to Figure 9.24. In the originating signalling node, the routing entity on the MTP level is called Signalling Route Set (SRS). SRS is the collection of the signalling routes through which a certain SPC can be reached. The signalling route is in practise the same as the SLS, but the difference here is that SLS is not "aware" of the STP facility, but the signalling route is.

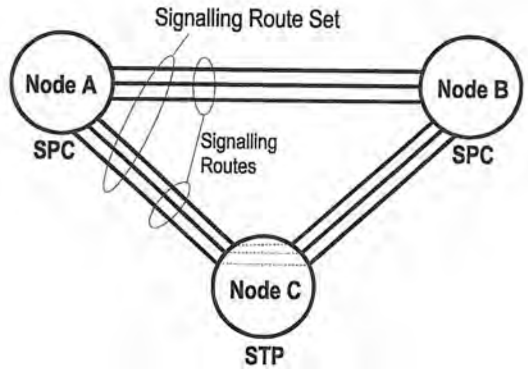


Figure 9.24. SS7 signalling route set

The STP nodes are capable of routing signalling messages only within a single SPC addressing space, which is typically administered by a single operator. In order to make it possible to route signalling across network boundaries the SCCP protocol is needed. SCCP uses global title addressing, which makes it possible, for example, for an VLR within a visited network to reach the HLR in the subscriber's home network by using the GT of the HLR for addressing.

9.3.5. IP Option for Signalling Transport

As described in the content of GTP-U above the user data within the PS domain backbone network and across the Iu–PS interface is transported with the IP-based protocol stack. Besides these two obvious cases the IP-based transport started to gain more momentum towards the end of the 3GPP R99 specification campaign. In order to create a basis for full-scale utilization of IP internetworking technology as a common transport for the future evolution of UMTS networks, it became necessary to study the IP option for signalling transport as well (Figure 9.25).

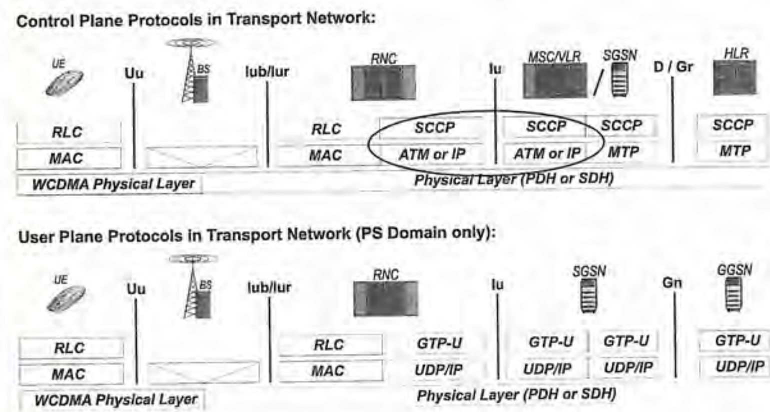


Figure 9.25. Transport network – signalling transport in UTRAN

In order to maintain the harmonization between the IP transport and ATM transport options both stacks should provide the same transport service. Therefore it was decided to "surface" both stacks with the SCCP protocol entity, which then always delivers a harmonized signalling transport service. The detailed structure of the resulting protocol stack is illustrated in Figure 9.26.

The left hand side of the stack has already been discussed in association with the ATM protocols above.

The adaptation of the SCCP protocol on top of the IP protocol stack was achieved by using two convergence protocols, which are developed and standardized by the SIGTRAN Working Group in IETF. These protocols and the corresponding IETF documents are

- Stream Control Transport Protocol (SCTP), RFC 2960
- MTP3 User Adaptation Layer (M3UA) (Internet draft document)

The purpose of these two sublayers is to provide the SCCP layer with an illusion of sitting on top of the regular MTP3 service as it has been in the telecommunication signalling networks for more than two decades.

In the 3GPP R99 transport network the SCTP protocol entities can be found in RNC and

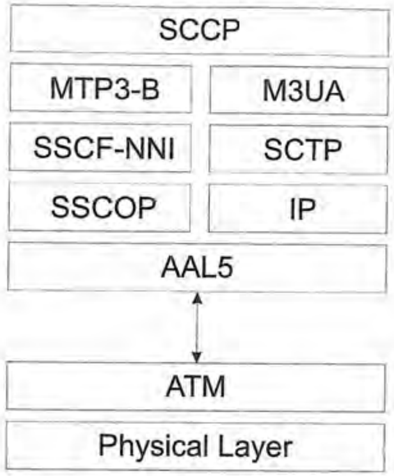


Figure 9.26. Two options for signalling transport stack

SGSN only. This transport option is available in the Iu–PS interface for both control plane and user plane as was targeted from the IP-evolution point of view.

SCTP assumes that it is running over an IPv4 or IPv6 network. Even more important is that SCTP assumes it is running over a well-engineered IP network. This, in practice, means that there is a diverse routing network underneath so as to avoid a single point of failure. It takes into consideration multi-homed endpoints, which are endpoints with more than one IP address/port number tuple, for additional reliability. Further more, it provides a MTU discovery function, to determine the MTU size of the data path, to avoid IP level fragmentation.

The purpose of the SCTP protocol is to provide a robust and reliable transport signalling bearer. To achieve this SCTP provides appropriate congestion control procedures, fast retransmit in the case of message loss and enhanced reliability. It also provides additional security against blind attacks, which will be used to increase security in connecting together UMTS networks of different operators.

The operational environment of the M3UA protocol is the same as that of the SCTP protocol. It is running on top of the SCTP protocol, which provides a reliable transport bearer to M3UA.

The purpose of the M3UA protocol is to support SCCP signalling, so that the SCCP lower interface does not need to be modified. M3UA needs to manage the use of SCTP streams and to meet equal performance as its SS7 counterpart, the MTP3b protocol, while running on top of an IP stack. M3UA provides mapping of SS7 addressing to IP addressing. It provides failover support as well as the ability to loadshare among endpoints.

M3UA has two architectural modes: signalling gateway to IP Signalling Point (IPSP) and IPSP-to-IPSP. It is assumed that the IPSP-to-IPSP is the most likely model and the simplest as well.

9.4. Radio Network Protocols

In the UMTS internetwork protocol architecture (see Figure 9.4) the radio network protocols compose the next layer on top of the generic transport network protocols discussed above. The radio network extends from UE across the whole UTRAN and terminates at the Iu edge nodes of the CN. The protocols in the radio network are needed to control the establishment, maintenance and release of radio access bearers and to transfer user data between UE and CN along those radio access bearers.

9.4.1. Radio Network Control Plane

The control plane protocols in the radio network layer execute all control needed for management of radio access bearers. According to the client–server approach applied in the overall design of the UTRAN control plane these protocols implement the radio resource management decisions and maintain the RABs for all UEs even when they are moving within the UTRAN area.

9.4.1.1. RANAP – Radio Access Network Application Protocol

The Iu interface connects the UMTS radio access network and the CN together. The same Iu interface is used to connect both CS service and PS service domains to the UTRAN (see Figure 9.27). The Iu interface has been designed to support the independent evolution of the UTRAN and CN technologies. In addition to that, it was made possible to develop independently CS and PS service domains inside the CN and still use the same Iu interface to connect the domains to the UTRAN.

From a UTRAN perspective it is desirable that the connections to the CS and PS service domains are as similar as possible. That is why a single signalling protocol between the

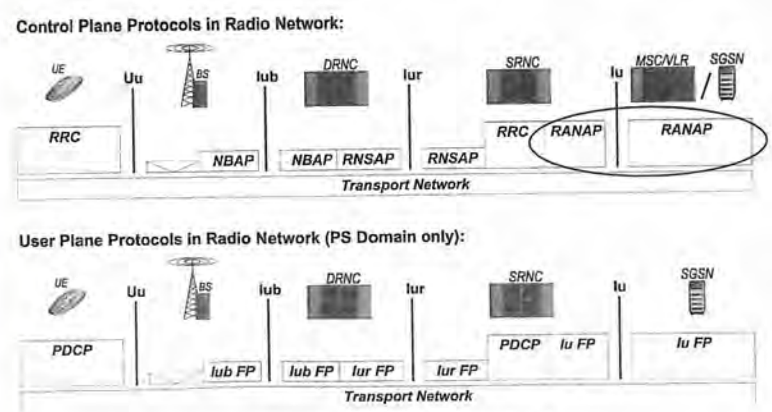


Figure 9.27. Radio network – RANAP

UTRAN and CN was defined. This protocol is called Radio Access Network Application Part (RANAP) and it is defined in the 3GPP specification TS 25.413. Both CS and PS domains use the RANAP protocol to access the services provided by the UTRAN.

RANAP is the protocol that controls the resources in the Iu interface. One RANAP entity resides in the RNC and the other peer entity in the MSC or the SGSN.

RANAP is located on top of the Iu signalling transport layers. RANAP uses the signalling transport service to transfer RANAP messages over the Iu interface. In 3GPP R99 the transport layers in the Iu interface are comprised of an SS7 protocol stack over ATM or IP over ATM.

The SS7 stack consists of several protocol layers but the topmost protocol is always SCCP. In spite of the two possible transport options (SS7 or IP) it is always an SCCP protocol entity, which offers RANAP the interface to the services of the transport protocol stack as was described in Section 9.3.5

RANAP has certain requirements for the transport layers below. Basically it is assumed that each message (PDU), which RANAP sends to the peer entity will reach the destination without any errors. In other words it is the responsibility of transport protocol stack to provide reliable data transfer across the Iu interface for the RANAP PDUs.

Each PDU of the dedicated control services should be sent on a unique signalling connection. In the Iu interface, signalling connection is realised by an Iu signalling bearer. The transport layers shall provide RANAP the means to dynamically establish and release signalling bearers for the Iu interface when RANAP requests it. Each active UE shall have its own Iu signalling bearer. It is the responsibility of the transport layers to maintain the bearers. If the Iu signalling bearer connection breaks for some reason, SCCP informs RANAP about it.

RANAP offers services to the upper layer protocols: on the MSC/VLR and SGSN side the non-access stratum protocols make use of the RANAP services. By interworking with the RRC protocol in RNC, RANAP contributes to the transfer of signalling messages of these upper layer protocols between the UE and CN.

RANAP services are divided into two groups, general control services and dedicated control services. General control services are related to the whole Iu interface between the RNC and CN. Dedicated control services support the separation of each UE in the Iu interface: they are always related to a single UE. The majority of RANAP services are dedicated services. All the RANAP messages of the dedicated control services are sent on a dedicated connection. This connection is called a signalling connection.

The overall RAB management is one of the main services RANAP offers. RANAP provides the means for the CN to control the establishment, modification, and release of the RABs between the UE and CN. Related to that, RANAP supports UE mobility by transferring RAB to a new RNS when the UE moves from the area of the serving RNS to another. This service is called SRNS relocation. Controlling the security mode in the UTRAN is one of the RANAP services as well. All the above-mentioned services are dedicated control services.

The general control services are needed only in exceptional situations. RANAP provides, for example, a means to control the overload in the Iu interface if the amount of the user traffic grows too high. In the case of a fatal failure in either end of the Iu interface RANAP offers a reset service. That initialises the whole Iu interface and clears all the active connections.

The detailed specification for the RANAP protocol is in the 3GPP specification TS 25.413.

9.4.1.2. RNSAP – Radio Network Subsystem Application Protocol

RNSAP is a control plane protocol at the radio network layer (Figure 9.28). It provides control signalling exchange across the Iur interface. RNSAP is run by two RNCs, one of which takes the role of Serving RNC (SRNC) and the other acts as a Drifting RNC (DRNC). SRNC is the one having the RANAP signalling connection to the CN.

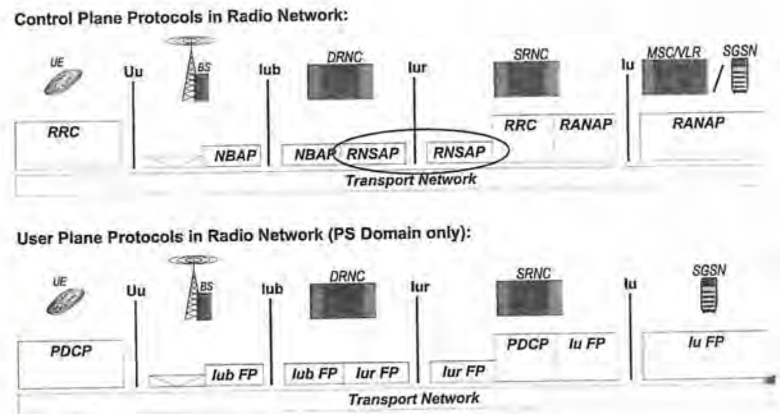


Figure 9.28. Radio network – RNSAP

The RNSAP protocol operates always on top of the SCCP protocol. Like in the case of the RANAP protocol there are two signalling transport options below SCCP: either ATM with an AAL5 adaptation layer or IP-based transport, where SCTP and M3UA protocols are used as the adaptation layer as described in Section 9.3.5. The SCCP as a signalling transport provides two different modes for RNSAP: connection-oriented and connectionless data transfer service.

RNSAP is responsible for bearer management signalling across the Iur interface. RNSAP is used for setting up radio links and allowing the SRNC to control those radio links using dedicated resources in a DRNS.

The RNSAP protocol uses one signalling connection per DRNC and UE where a UE may have one or more active radio links for the transfer of layer 3 messages.

The information transferred over the Iur interface can be categorised as follows:

- Radio and mobility control signalling: the Iur interface provides the capability to support radio interface mobility between RNSs for the UEs having a connection with UTRAN. This capability includes the support of handover, radio resource control and synchronisation between RNSs.
- Iub/Iur DCH data streams: the Iur interface provides the means for transport of uplink and downlink Iub/Iur DCH frames carrying user data and control information between SRNC and remote BS, via the DRNC.
- Iur DSCH data streams: an Iur DSCH data stream corresponds to the data carried on one

DSCH transport channel for one UE. A UE may have multiple Iur DSCH data streams. In addition, the interface provides a means to the SRNC for queue reporting and for the DRNC to allocate capacity to the SRNC.

- Iur RACH/CPCH data streams: in WCDMA FDD variant only.
- Iur FACH data streams.

The main UTRAN control functions supported by the RNSAP protocol exchange are:

- Transport network management.
- Traffic management of common transport channels, e.g. paging.
- Traffic management of dedicated transport channels, which includes, e.g. radio link set-up, addition and deletion as well as measurement reporting.
- Traffic management of downlink shared transport channels, which includes, e.g. radio link set-up, addition and deletion as well as capacity allocation.
- Measurement reporting for common and dedicated measurement objects.
- SRNS relocation, this function co-ordinates the activities when the serving RNS role is to be taken over by another RNS.

The RNSAP protocol participates in the radio link set-up procedure when it is used for establishing the necessary resources in the DRNC for one or more radio links. The connection-oriented service of the signalling bearer shall be established in conjunction with this procedure.

Other UTRAN-wide procedures, which may require protocol activities from the RNSAP entities are soft and hard handovers, cell update, URA update and paging as well as downlink power control actions and the RRC connection release and re-establishment whenever it needs to be performed over the Iur interface.

The complete specification of the RNSAP protocol is in the 3GPP specification TS 25.423.

9.4.1.3. NBAP – Node B Application Protocol

NBAP is a radio network layer protocol, which maintains control plane signalling across the Iub interface and thus controls resources in the Iub interface and provides a means for BS and RNC to communicate (Figure 9.29). The 3GPP specification TS 25.433 defines functionality of the NBAP protocol. One peer entity of NBAP resides in BS and for each BS the other entity resides in that RNC, which controls the BS, i.e. in the CRNC.

NBAP resides on top of the Iub transport layers and uses their services to transfer NBAP messages over Iub interface. The 3GPP R99 choice of underlying transport technology used to carry NBAP signalling over the Iub interface is ATM. The ATM connection together with the necessary ATM adaptation layers provides a signalling bearer for NBAP as defined in the 3GPP specification TS 25.432. The signalling bearer for NBAP provides reliable point-to-point connection. This means that NBAP assumes that each message NBAP sends to its peer entity will reach its destination without any errors. There may be multiple point-to-point links for NBAP signalling at the Iub interface. Since the Iub interface conforms to the generic model for UTRAN terrestrial interfaces, the specification of NBAP is also done so that functionality of NBAP does not depend on the transport layers below. New transport layers can be used with NBAP without the need to redefine the functionality of the protocol itself.

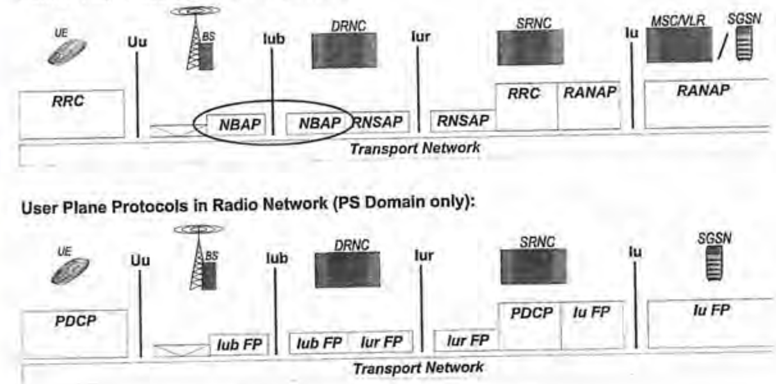


Figure 9.29. Radio network – NBAP

BS does not make any decision of its own about the radio resource or RAB management but rather obeys commands from the RNC. NBAP offers its services to the control units of RNC and BS. NBAP also has an interface with the RRC protocol to provide the RRC entity in BS with information, which should be mapped to the BCCH logical channel. NBAP has no other interfaces to radio network layer protocols of UTRAN and all interactions with them are carried out via the control unit of RNC.

Figure 9.30 illustrates the operational environment of the NBAP protocol and all the interfaces of NBAP are depicted as bold black arrows on the figure.

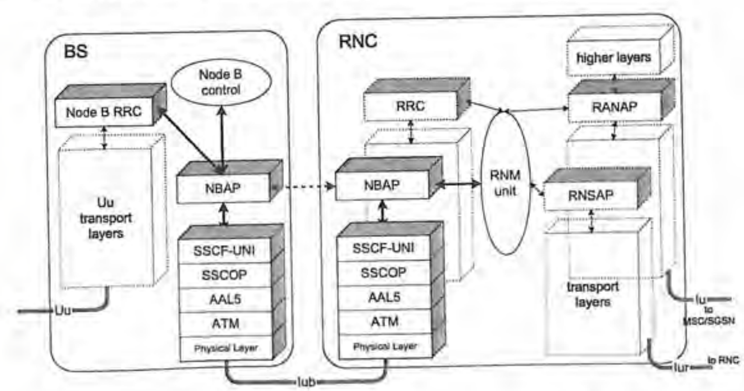


Figure 9.30. Operational environment of NBAP

All the resources physically implemented in BS logically belong to and are controlled by RNC. Therefore, the physical resources of a BS are seen as logical resources by the CRNC. This RNC performs the control on the level of logical resources, while the physical implementation of resources and control procedures in the BS is left unspecified. Physical procedures performed in the BS change the conditions of the logical resources owned by RNC and this requires some information exchange between RNC and the BS to keep the RNC aware of the changes. Such information exchange is referred as Logical Operation and Maintenance (Logical O&M) of the BS. Logical O&M constitutes an integral part of NBAP signalling.

Another important responsibility of NBAP is to establish and maintain a control plane connection over the Iub interface, to initiate set-up and release of dedicated user plane connections across the Iub interface and command the BS to activate resources for new radio links over the Uu interface. All NBAP signalling functions are divided into common procedures and dedicated procedures.

The common NBAP procedures are not related to any particular UE but to common resources across the Iub interface. Signalling related to Logical O&M of the BS constitutes the great part of NBAP common procedures. Common NBAP includes procedures for configuration management of logical resources and procedures, which enable the BS to inform the RNC about the status of logical resources in the BS. Common NBAP also allows the RNC to initiate specific measurements in the BS and the latter to report the results of the measurements. Besides Logical O&M, common NBAP is used to deliver from RNC to BS the information to be transported on broadcast channel. Common NBAP also initiates the creation of a new UE context for each specific UE in BS by setting up the first radio link for that UE. Common NBAP maintains a control plane signalling connection across the Iub interface. For this reason there is always one signalling link for common NBAP at the Iub interface, which terminates at the common control port of the BS.

The dedicated NBAP procedures are always related to one specific UE, for which UE context already exists in BS. Dedicated NBAP includes procedures for management and supervision of existing radio links, whose UE is connected to the BS. These procedures allow the RNC to command the BS to establish or release some radio links for the UE context. Dedicated NBAP also provides the BS with ability to report failure or restoration of transmission on radio links. Dedicated procedures also include radio link reconfiguration management and support measurements on dedicated resources and corresponding power control activities, which allow the RNC to adjust downlink power level on the radio links. Dedicated NBAP signalling is carried out via one of the communication control ports of the BS logical model. There can be several dedicated signalling links for dedicated NBAP across the Iub interface.

The complete specification of the RNSAP protocol is in the 3GPP specification TS 25.433.

9.4.1.4. Radio Resource Control (RRC) Protocol

The RRC protocol is the key radio resource control protocol within UTRAN (Figure 9.31). It supports the RRM functionality discussed in Chapter 4 by coordinating the execution of resource control requests, which result from the decisions made by the RRM algorithms. With the help of the RRC protocol the effect of radio resource management decisions can be extended to all the UTRAN network elements affected by such decisions. The RRC protocol

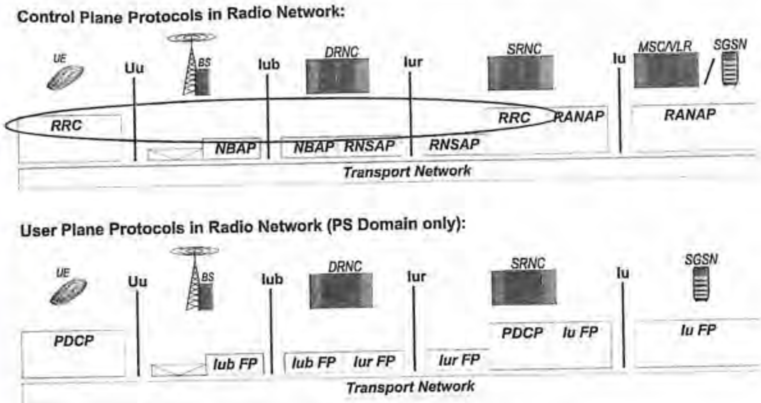


Figure 9.31. Radio network - (RRC) protocol

messages also carry in their payload all the signalling belonging to the non-access stratum protocols.

The RRC protocol operates between the UE and RNC. RRC protocol entities use the signalling bearers provided by the RLC sublayer to transport the signalling messages. All three modes of the RLC are available as different kinds of SAPs as shown in Figure 9.32. The

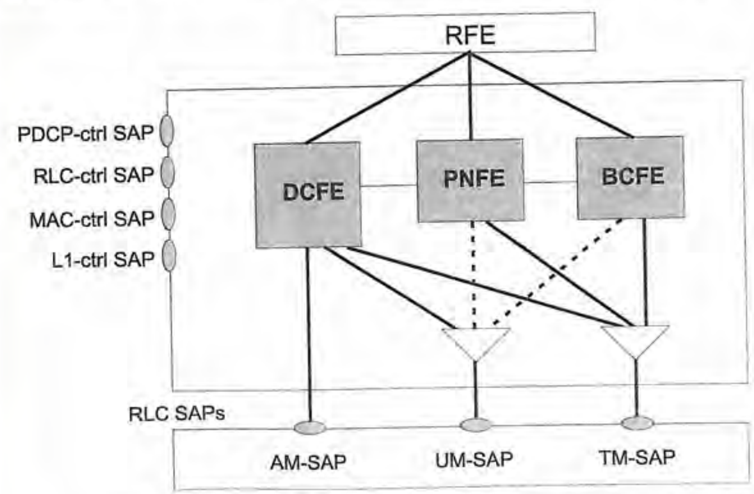


Figure 9.32. Logical structure of an RRC protocol entity in WCDMA-FDD

transparent mode TM-SAP is used, for example, whenever a UE has to communicate with an RNC prior to establishment of a full RRC connection (e.g. initial access or cell/URA update with a new cell). It is also used for frequently repeated messages, like paging, to avoid unnecessary overheads. The acknowledged mode AM-SAP is used for the control signalling specific to one UE whenever the reliability of the message exchange is required. The unacknowledged mode UM-SAP on the other hand is used to avoid the potential delay present in the acknowledged mode signalling. For example when releasing an RRC connection the release message is repeated many times (quick repeat function) via UM-SAP to increase the probability of the UE receiving it.

Figure 9.32 also shows the logical structure of an RRC protocol entity. The Dedicated Control Function Entity (DCFE) is used to handle all signalling specific to one UE. In SRNC there is one DCFE entity for each UE having an RRC connection with this RNC. The Paging and Notification Control Entity (PNFE) handles paging messages to the idle mode UEs. In a CRNC there is at least one PNFE entity for each cell to be controlled. The Broadcast Control Function Entity (BCFE) handles system information broadcasting on BCCH and FACH logical channels. There is at least one BCFE entity for each cell in a CRNC. Besides these functional entities a special routing function entity (RFE) is also modelled on top of Figure 9.32. Its task is to route the non-access stratum messages to different MM/CM entities on the UE side and to different CN domains on the RNC side.

As the key executor of the radio resource allocation the RRC entities in the UE and RNC have control over the L1, MAC and RLC entities on their side. In order to execute the control commands on these lower layer protocol entities special control SAPs are available to the RRC protocol entity as shown in Figure 9.32. These control SAPs are also used for reporting measurements and exceptional conditions detected by the lower layers.

The major function of the RRC protocol is to control the radio bearers, transport channels and physical channels. This is done by set-up, reconfiguration and release of different kinds of radio bearers. Before such actions can take place the RRC protocol communication itself must be initiated by using a number of (minimum four) signalling radio bearers. The resulting signalling connection together with any subsequently established other bearers is called *RRC connection*. During the RRC connection, set-up, reconfiguration and release of other radio bearers for user plane traffic may be executed by exchanging commands and status information between the peer RRC entities over the signalling radio bearers. The RRC connection will continue to exist until all user plane bearers have been released and the RRC connection between UE and RNC is explicitly released.

The security mechanisms applied on the radio bearers are activated and deactivated under the control of the RRC protocol. Besides the activation of confidentiality protection by ciphering, the RRC protocol can also guarantee the integrity of all higher layer signalling messages as well as most of its own signalling messages. This property, which is discussed more thoroughly in Chapter 8, is achieved by attaching to every RRC PDU a 32 bit message authentication code, which is calculated and verified by the RRC entities themselves. It ensures that the receiving RRC entity can verify that signalling data has not been modified and that the data has really originated from the claimed peer entity. Also the operation to change the keys used by these functions is performed by the RRC protocol under the instructions coming originally from the CN.

The UE mobility management at the UTRAN level is controlled by RRC signalling. Such

mobility management functions executed by the RRC protocol are cell update, URA update and active set update, which have been discussed in Chapter 4.

Control and reporting of radio measurements is taken care of by the RRC protocol. A total of seven different categories of measurements can be activated at the UE, among them also the measurements for UE positioning. Each measurement can be controlled and reported independently from the other measurements.

Besides the above-mentioned tasks the RRC protocol also controls the broadcasting of system information and paging of UEs and exchange parameters for power control purposes.

As a protocol RRC is fairly complex. Altogether it has some 40 different procedures and more than 60 different kinds of PDU.

The RRC protocol is specified in the 3GPP specification TS 25.331 and a more detailed description can be found in Holma and Toskala (2001).

9.4.2. Radio Network User Plane

9.4.2.1. PDCP – Packet Data Convergence Protocol

As the name suggests the Packet Data Convergence Protocol (PDCP) is designed to make the WCDMA radio protocols suitable for carrying the most common user-to-user packet data protocols, TCP/IP (Figure 9.33). Therefore PDCP protocol entities can be found on both sides of the WCDMA radio interface: at the RNC and UE. The key functionality of the PDCP protocol is to compress the headers of the payload protocols, which – if sent without compression – would waste the invaluable radio link capacity. The only IP header compression specified for the 3GPP R99 is RFC2507 defined by the IETF.

In the 3GPP radio interface protocol model PDCP belongs to the radio link layer (L2), the topmost sublayer of it is specially designed for user-plane radio bearers carrying packet data.

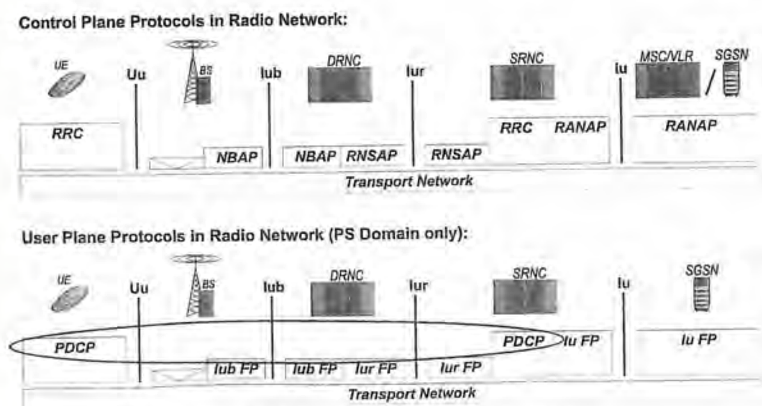


Figure 9.33. Radio network user plane – PDCP

The PCDP protocol makes use of the RLC layer services in three different modes. Conceptually those services are available at three different Service Access Points (SAPs):

- Data transfer is acknowledged (AM-SAP)
- Data transfer is unacknowledged (UM-SAP)
- Data transfer in transparent mode (TM-SAP)

The PDCP protocol assumes that the segmentation and reassembly is taken care of by the RLC layer below. When lossless SRNS relocation is supported, then PDCP uses an AM-SAP and assumes in-sequence delivery of RLC SDUs.

PDCP uses the AM-SAP whenever reliable data transfer is requested. Then the RLC layer is only allowed to discard some of the RLC SDUs, if configured in such a way, but is expected to be able to indicate the PDCP of each the discarded SDUs both in the transmitter and the receiver. These indications are made necessary by the fact, that – as a compression protocol – the PDCP avoids sending explicit sequence numbers over the radio, but instead makes use of a virtual sequence numbering scheme. This makes it also necessary to let the PDCP know about possible resetting of the RLC connection and radio bearer reconfigurations, especially when lossless SRNS relocation is carried out.

For TM-SAP the size of the PDCP PDUs is expected to be fixed, which means very strict requirements either for the PDU size of the upper layer (e.g. IP protocol) or for the applied IP header compression. In a generic case neither the upper layers in the PS domain nor the currently specified IP header compression method can satisfy these requirements and therefore TM-SAP should be used very carefully and in well studied cases only.

The services provided by the PDCP protocol to the layers above are simply as follows:

- Delivery of packets (SDUs) either in acknowledged, unacknowledged or transparent mode.
- Radio bearer control with set-up, reconfiguration and release of a radio bearer in the PS domain.

The PDCP protocol entity running at the RNC has to interwork with the UTRAN frame protocols in order to relay the data packets in the downlink direction from Iu to Iub/Iur and in the uplink direction from Iub/Iur towards Iu. In normal cases this is achieved simply by copying packets from one interface to the other and performing the specified header compression/decompression on the way. Things get more complicated whenever an intersystem handover or lossless SRNS relocation needs to be performed. The PDCP protocol is therefore equipped with special interworking rules with the Iu frame protocols and the RRC protocol which ensure that packets are not lost or duplicated and that the packet sequence numbering can be restored after the handover or relocation at the target system.

In the 3GPP release 4/5 another header compression algorithm called Robust Header Compression (ROHC) as defined in the IETF RFC3095 will be added to the PDCP specification.

The complete specification of the PDCP protocol is in the 3GPP specification TS 25.323 and more information can also be found in Holma and Toskala (2001).

9.4.2.2. UTRAN Frame Protocols

The UTRAN frame protocols are the radio network layer user plane protocols, which carry the UMTS user data over the common UTRAN transport network (Figure 9.34). They are active across the UTRAN interfaces Iu, Iur and Iub. They use the transport network service, which is somewhat interface-specific.

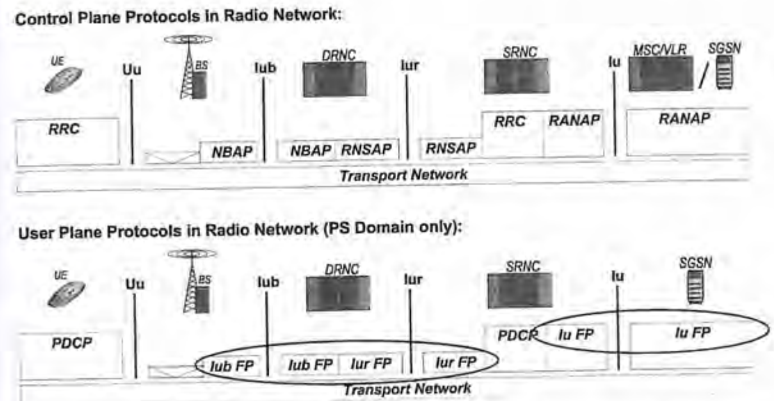


Figure 9.34. Radio network user plane – UTRAN frame protocols

Frame Protocol at Iu

The frame handling protocol in the Iu interface has been designed to meet the needs of both the CS domain and PS domain user data traffic. The transport services for these two domains are different and therefore two modes of Iu frame protocol are used:

Transparent mode, which is used on such radio access bearers, which do not require any special protocol activities, like framing. Therefore the user data in this mode is simply passed through to the transport network stack without adding any protocol header information. Transparent mode hence becomes a null protocol. It is used, e.g. for carrying non-real time data packets in their plain GTP-U format from RNC to SGSN and vice versa.

Support mode, which takes the form of a real protocol with control PDUs and framing of user data. These protocol formats are applied on radio access bearers, which carry, e.g. AMR coded speech data in the CS domain. Therefore the support mode control PDUs can take care of rate control and time alignment, which are needed to support real-time speech transport. In this case the support mode protocol uses directly the AAL2 service provided by the underlying transport network.

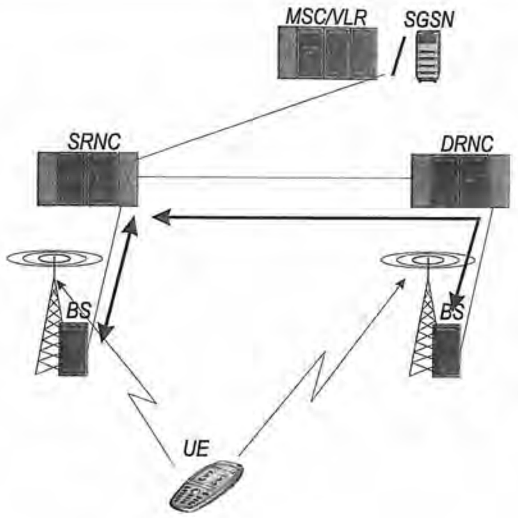


Figure 9.35. Frame protocol for Iub/Iur dedicated channels

Frame Protocol for Dedicated Channels at Iur and Iub

There is a common frame protocol for dedicated channels over the Iub and Iur as illustrated in Figure 9.35. With this protocol the SRNC can exchange user data frames with UEs being

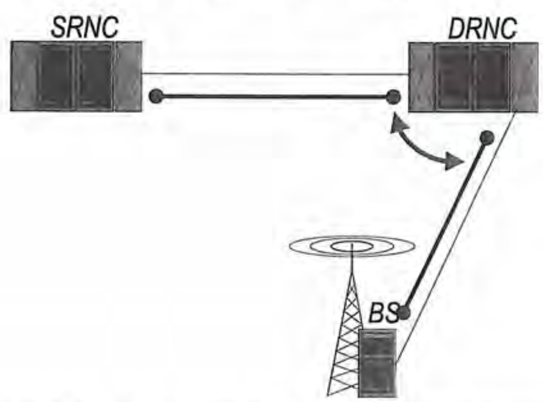


Figure 9.36. Frame protocols for Iub/Iub common transport channels

serviced by its own as well as remote BSs. Besides the normal uplink and downlink data transfer, this frame protocol is capable of performing many control functions. Such protocol functions deal with timing adjustment and synchronisation. Also outer loop power control commands can be sent to UEs in dedicated channel frame protocol messages as well as other updates to the radio interface parameters, which need to be made known to the UEs.

Frame Protocols for Common Transport Channels at Iur and Iub

For sending and receiving data frames over common transport channels there are two different frame handling protocols: one for the Iub and another one for the Iur. Therefore, interworking between these protocols is done at the DRNC as illustrated in Figure 9.36.

Besides the uplink and downlink data transfer, which is to some extent specific to different transport channel types, these frame protocols also carry out control tasks. On the Iur frame handling flow control based on credits is applied. In the Iub frame handling synchronisation and timing adjustment needs are taken care of.

9.5. System Network Protocols

Once the radio network protocols make it possible to communicate across the UTRAN subnetwork by maintaining the communication path to the mobile terminals, it is the system network protocols, which create the communication services to the users of those terminals. The system (network) protocols operate on top of the radio network and within the UMTS CN itself.

9.5.1. Non-Access Stratum Protocols

The non-access stratum refers to the group of control plane protocols, which control the communication between UEs and CN. The common name "non-access stratum" refers to the fact that these protocols are carried transparently through the radio access network.

The protocols in this group belong to two sublayers of the system network as shown in Figure 9.3. The MM sublayer operates on top of the signalling connection provided by the radio network. Besides its own functionality, the MM sublayer acts as a carrier to the CM sublayer protocols.

9.5.1.1. MM Protocol for the CS Domain

MM protocol operates between UE and MSC/VLR (Figure 9.37). This control plane protocol provides basic signalling mechanisms for controlling mobility management and authentication functions between UEs and the CN CS domain. The MM protocol is supported by UEs operating either in the PS/CS operation mode or in the CS-only operation mode.

The MM protocol makes use of the signalling connection provided by the radio network layer protocols. This signalling connection is sometimes referred to as the radio resource (RR) connection, which consists of the RRC protocol connection over the signalling radio bearer and the RANAP connection over the Iu signalling bearer. Before any MM protocol activity can take place the signalling connection must first be established. Inside the CN the MM protocol entity on the MSC/VLR side has to interwork with the MAP protocol in order to maintain the location registration in the HLR.

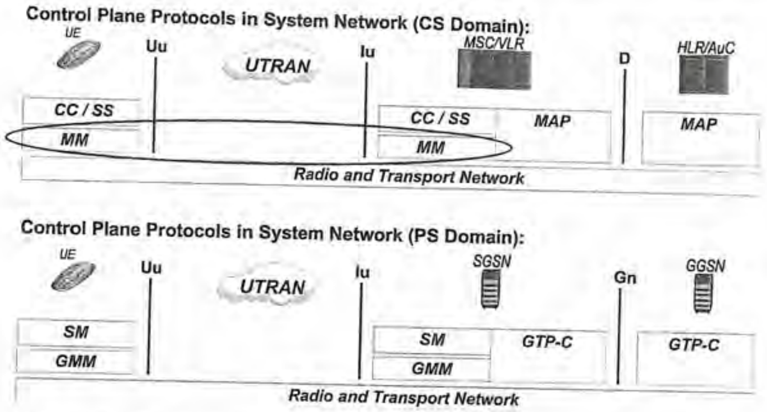


Figure 9.37. System network non-access stratum – MM protocol for the CS domain

The MM protocol controls the MM state of the UE and the corresponding peer entity on the network side. The major states of these protocol entities indicate the location update status of the UE, which has a major impact on its ability to use the services of the CM sublayer. This is also why the MM protocol is used for carrying messages of the CM sublayer protocols between UE and the MSC/VLR.

The MM protocol has three basic categories of procedures:

- MM connection procedures
- MM common procedures
- MM specific procedures

The *MM connection procedures* are used to establish and release MM connections as well as for transferring the CM sublayer messages. The concept of MM connection is established without exchange of any special MM protocol initiation messages. Instead, an MM connection is established and released implicitly on request from a CM entity.

When the MM connection establishment is initiated by the UE, the MM protocol entity issues a CM SERVICE REQUEST, which contains the following parameters

- Identity of the UE, e.g. TMSI;
- Mobile station classmark 2;
- Ciphering key sequence number; and
- CM service type identifying the requested type of transaction, which can be, e.g. mobile originating call establishment, emergency call establishment, short message service or supplementary service activation.

The MM protocol entity receiving this request on the CN side will analyse these parameters and decide on approval and continuation of the service request.

On the other hand the request to establish an MM connection may come from the CM entity

on the network side. This is then accomplished by making the UE to establish MM connection via the paging procedure.

After the MM connection has been established, it can be used by the CM sublayer entity for information transfer. According to the protocol architecture shown in Figure 9.3, each CM entity will have its own MM connection. These different MM connections are identified by the Protocol Discriminator (PD) and, additionally, by the Transaction Identifier (TI).

An established MM connection can be released by the local CM entity. The release of the connection will then be done locally in the MM sublayer, i.e. no MM messages are sent over the radio interface for this purpose.

The *MM common procedures* may be initiated at any time while MM connections are active. These procedures are used mainly for security functions when the identity and location of the UE is known by the CN and the MM connection exists from the CN to the UE. The mandatory MM common procedures are:

- *TMSI reallocation procedure*, which is used to provide confidentiality of the user identity, i.e. to protect a user against being identified and located by an intruder. In this procedure a temporary random identification (TMSI) is allocated by the MSC/VLR for use within radio interface procedures instead of the permanent, and therefore traceable, IMSI identification. The procedure is initiated by the CN and is usually performed in the ciphered mode.
- *Authentication procedure*, which permits the CN to check whether the identity provided by the UE is acceptable or not, and provide parameters enabling the USIM to calculate new ciphering and integrity keys. In a UMTS authentication challenge, the authentication procedure is extended to allow the UE to check the authenticity of the CN (see Chapter 8). This allows, for instance, detection of a false base station. The authentication procedure is always initiated by the CN.
- *Identification procedure*, which is used by the CN to request the UE to provide specific identification parameters to the network like IMSI and IMEI.
- *IMSI detach procedure*is initiated by the UE when the UE is deactivated or if the USIM is detached from the ME. Each UMTS network can broadcast a flag within the system information block to indicate whether the detach procedure is required by UEs. When receiving an MM IMSI DETACH INDICATION message, the network may set an inactive indication for this IMSI. No response is returned to the UE. After this the network shall release locally any ongoing MM connections, and start the normal signalling connection release procedure.
- *Abort procedure* can be used by the CN to abort any on-going MM connection establishment or already established MM connection due to a network failure or having found an illegal ME.

The *MM specific procedures* handle messages used in location update procedures. An example message flow of a location update procedure is described in Chapter 10. There are three variants of the basic location update procedure:

- *Normal location updating procedure* for the situation where the UE has detected that its location area is changed and it is necessary to inform the CN about the new location. The detection is done by checking the Location Area Identifier (LAI) value broadcasted by the network against the current LAI value stored in the USIM. The normal location updating procedure shall also be started if the network indicates that the UE is unknown in the VLR as a response to service request. The UE has to maintain a list of "forbidden location areas for roaming" on the USIM based on failed location update requests.

- *Periodic location updating procedure* may be used to notify periodically the availability and location of the UE to the CN according to a timer in the UE. The timeout value for this timer is taken from the broadcasted system information block.
- *IMSI attach procedure* is performed when the UE enters (or is switched on in) the network in the same location area, where it was previously detached, i.e. when the LAI value broadcasted in the current cell is detected to be equal to the LAI value on the USIM.

The common characteristic feature of the MM specific procedures is that they can only be initiated when no other MM specific procedure is running or when no MM connection exists. Notice also that any of the MM common procedures described above (except IMSI detach) may be initiated during a MM specific procedure.

Since the MM protocol is a direct extension of its GSM predecessor, a reader interested in more details about it may consult any of the GSM textbooks listed in the Bibliography. The complete specification of the MM protocol together with the UMTS specific details is found in 3GPP specification TS 24.008.

9.5.1.2. GPRS Mobility Management for PS Domain (GMM)

GMM operates between the UE and SGSN (Figure 9.38). The name GPRS MM is used because the protocol is basically evolved from the corresponding protocol for packet switched services in the GSM/GPRS system. This control plane protocol provides basic signalling mechanisms for controlling MM and authentication functions between UEs and the CN PS domain. The GMM protocol is supported by UEs operating either in the PS/CS operation mode or in the PS operation mode.

Like the MM protocol, the GMM protocol makes use of a signalling connection between

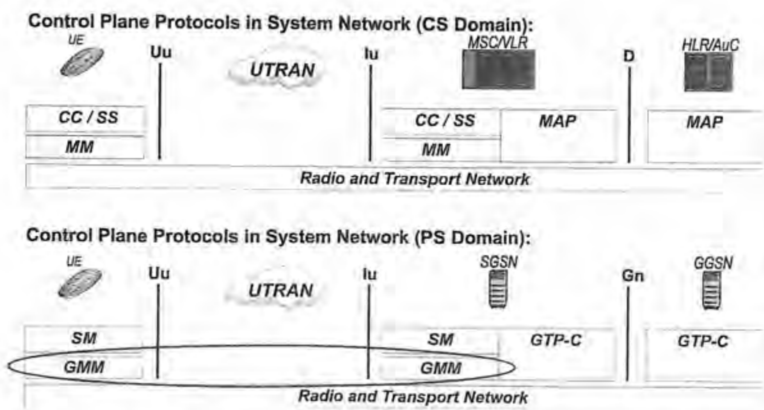


Figure 9.38. System network non-access stratum – MM protocol for the PS domain

UE and SGSN. This signalling connection – which is sometimes called PS signalling connection – is provided by the radio network layer protocols using RRC over the signalling radio bearer and RANAP over the Iu signalling bearer. The GMM protocol entity on the SGSN side has to interwork with the GTP-C protocol entity for location information exchange.

It is possible that a UE operating in PS/CS operation mode has signalling connections both to the MSC/VLR and to SGSN simultaneously. For optimisation reasons it is defined that in these circumstances the GMM protocol entities in both UE and SGSN may perform so called combined procedures. The idea of the combined GMM procedures is to avoid sending both MM and GMM messages when the CN PS domain can inform the CN CS domain about such combined activities inside the CN. The optional Gs interface is used to take this action on the CN side.

From the packet data services point of view it is important to distinguish between the state, when the UE together with its identity and location is known by the CN PS domain and the state, when it is not known. The importance comes from the fact that the packet user data is protected by encryption which makes use of user-specific keys. This separation is highlighted by the *GMM context* concept. When the user is known to the CN PS domain network the GMM context exists, otherwise it does not exist.

GMM procedures for explicit establishment and release of the GMM context are:

- *GPRS attach* and *combined GPRS attach* procedures are used to indicate the user identity (typically P-TMSI) and current routing area to the network and to establish a GMM context for this user. A typical situation to execute this procedure is when the packet mode UE is switched on or USIM is inserted into it. In the case of the combined GPRS attach procedure the activities for normal IMSI attach are carried out.
- *GPRS detach* or *combined GPRS detach* procedures are invoked when the UE is switched off or the USIM is removed from the UE or a network failure has occurred. The procedures cause the UE to be marked as inactive in the CN PS domain only or in the combined GPRS detach procedure also for the CN CS domain.

The following GMM procedures are used for location management when a GMM context already exists:

- Normal *routing area updating* and *combined routing area updating* procedures are performed when the UE recognises that the routing area is changed to inform the CN PS domain about the event. In the combined routing area updating the location area is also included for the CN CS domain.
- *Periodic routing area updating* procedures may be used to notify periodically the availability of the UE to the CN according to a timer in the UE.

The following GMM common procedures for security purposes can be initiated when GMM context for the user has been established.

- A *P-TMSI re-allocation* procedure provides a temporary random identification, packet-TMSI (P-TMSI) for use within radio interface procedures instead of the permanent, and therefore traceable, IMSI identification. Normally the P-TMSI re-allocation procedure is performed in conjunction with another GMM procedure like routing area updating.
- A *GPRS authentication and ciphering* procedure, initiated always by the CN, permits the CN to check whether the identity provided by the UE is acceptable. It also provides parameters enabling the USIM to calculate UMTS ciphering and integrity keys and

permits the UE to authenticate the CN. See Chapter 8 for more details about these security functions.

- A *GPRS identification procedure* is used by the CN PS domain to request the UE to provide specific identification parameters to the network like IMSI and IMEI.

One more GMM procedure is needed to provide services to the CM sublayer on top of the GMM protocol:

- A *service request* procedure, which is initiated by the UE, is used to establish a secure connection to the network and/or to request the bearer establishment for sending data. The procedure is typically used when the UE has such a signalling message (e.g. SM or paging response message) or user data packet, that requires security protection, to be sent.

More details about the GMM protocol can be found in the 3GPP specification TS 24.008.

9.5.1.3. Call Control (CC) Protocol

The CC protocol provides basic signalling mechanisms for establishing and releasing circuit switched services (e.g. voice call or multimedia call) (Figure 9.39). The CC protocol operates between an MSC/VLR and a UE operating in the CS/PS or CS-only mode. The CC protocol uses the connection service provided by the MM sublayer to carry the CC protocol messages.

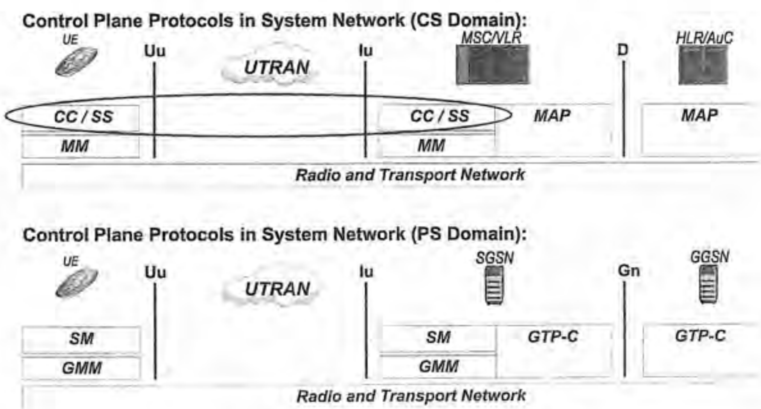


Figure 9.39. System network non-access stratum – CC protocol

The normal circuit switched voice call is controlled pretty much like in the ISDN telephony. Therefore the CC protocol entity in the GMSC has to interwork with the ISDN User Part (ISUP) protocol for establishment of a call path through an external ISDN network.

The UMTS circuit switched multimedia call is based on the 3G-324M. The 3G-324M is a variant of the ITU-T recommendation H.324. For details see the 3GPP specification TS 26.111.

The call establishment procedures establish the CC connection between the UE and the CN

and activate the voice/multimedia codec and interworking functions, if needed. The CC protocol entity in the MSC/VLR also interworks with the RANAP protocol entity in order to establish a RAB for the circuit switched call. This means, that the QoS requirements received by the CC protocol entity in the CC SETUP message have to be mapped onto a RAB request for the RANAP protocol entity in the same MSC/VLR. Only after the UTRAN has confirmed the assignment of a RAB, the call establishment can then proceed.

The established calls can be Mobile Originating (MO) or Mobile Terminating (MT). For multimedia circuit switched calls the normal call establishment procedure applies with some extra information elements in the CC protocol messages. The call release procedure, initiated by the UE or the CN release the resources for the call.

Call collisions as such cannot occur at the network, because any simultaneous MO or MT calls are dealt with separately assigned and different transaction identifiers in the CC protocol.

CC as a core network function was discussed in Chapter 5. An example of message flows for circuit switched call establishment and release are given in Chapter 10.

Since the CC protocol is a direct extension of its GSM predecessor, a reader interested in more details about it may consult any of the GSM textbooks listed in the Bibliography. The full specification of the CC protocol together with the UMTS specific details can be found in the 3GPP specification TS 24.008.

9.5.1.4. Session Management (SM) Protocol

The protocol entities running the SM protocol are located in the UE and SGSN (Figure 9.40). This protocol is the counterpart to the circuit switched CC protocol in the sense, that it is used to establish and release packet data sessions. To underline the possibility of using several packet data protocols within the UMTS network these sessions are called PDP contexts as discussed in Chapter 5. In practice the single most important packet data protocol to be supported is IP and especially the IPv6.

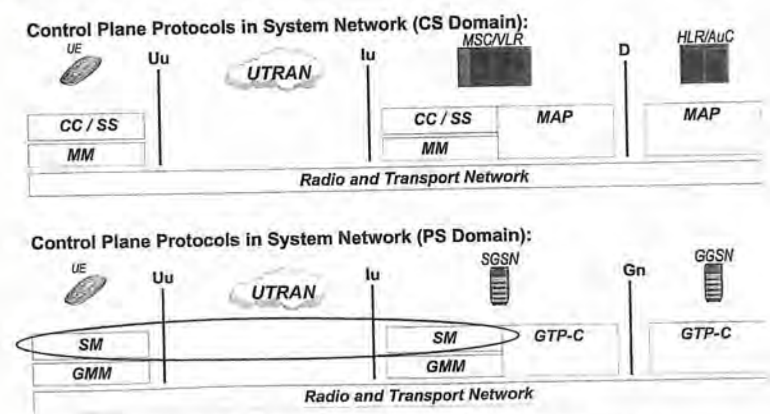


Figure 9.40. System network non-access stratum – SM protocol

The main purpose of the SM protocol is to support PDP context handling of the UE. In SGSN the SM protocol entity interworks with the GTP-C protocol entity on the Gn interface side in order to extend the PDP context management to the GGSN. The PDP context contains the necessary information for routing user plane data packets from the GGSN to the UE and vice versa. A UE may have more than one PDP context activated simultaneously in which case each of them is controlled by a SM protocol entity of its own.

The SM procedures require the existence of a GMM context for the identified user. If no GMM context for the user has been established, the MM sublayer has to initiate a GMM procedure to establish the needed GMM context (see GMM protocol above). After the GMM context establishment, the SM protocol uses services offered by the MM sublayer. Ongoing SM procedures are suspended during a GMM procedure execution.

The SM procedures can be initiated either by the GGSN or by the UE. Basic SM procedures are:

- The PDP context activation procedure, which establishes a new PDP context at the UE, SSGN and GGSN with necessary routing and QoS parameters for the session.
- The PDP context modification procedure is invoked when a change in the session parameters like QoS or traffic flow template is needed during the session. The modification may be initiated either by the UE or the GGSN.
- The PDP context deactivation procedure destroys PDP contexts in the UE and CN elements when the corresponding session is no longer needed.

A more detailed description on the SM protocol is given in the 3GPP specification TS 24.008.

9.5.1.5. Supplementary Services (SS) Protocol

Supplementary services are additional services defined to be used over circuit switched connections (Figure 9.41). This concept has been inherited from the 2G GSM system and

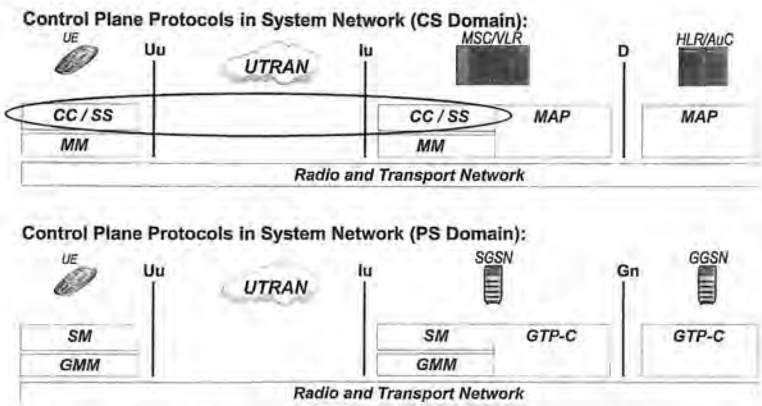


Figure 9.41. System network non-access stratum – SS protocol

therefore a UMTS terminal operating in the CS mode has the same ability for SS as a GSM mobile phone.

The 3GPP R99 implementation supports all the SS as far as circuit switched traffic is concerned. Some of the typical supplementary services are:

- Call forwarding services with different triggering conditions: no reply, busy, unconditional, not reachable.
- Call barring services for all incoming, all outgoing, all outgoing international or roaming calls.
- In-call services like call wait and call hold.

A complete list and description of SS, which are supported by the 3GPP R99 can be found in specification TS 24.080.

SS are controlled by the SS protocol, which is part of the CM sublayer. The SS protocol entities on the UE and MSC/VLR side communicate by using the MM connection. The communication has to do with activation and deactivation of the supplementary services. On the CN side the SS protocol interworks with the MAP-protocol, which is used whenever the supplementary service command requires, e.g. access to the subscriber information in the HLR.

9.5.2. Control Plane Between CN Nodes

Section 9.5.1 briefly introduced the control plane protocols in the non-access stratum between the UE and the CN. Despite the fact that the CN is divided into two functional domains, circuit and packet switched domains, these entities also have to perform some signalling functions with each other and especially with the registers located in the home network part of the CN.

MAP and CAP protocols have already been mentioned in Section 9.1.3 and an interested reader may find more detailed descriptions on these protocols in the GSM textbooks listed in the Bibliography. The 3G version of the MAP protocol is defined in the 3GPP specification TS 29.002.

9.5.2.1. GTP-C – GPRS Tunnelling Protocol for Control Plane

GTP-C belongs to the PS domain in the CN and is located in SGSN and GGSN (Figure 9.42). GTP-C is a control plane protocol specifying tunnel management and control procedures in order to allow SGSN and GGSN to provide user data packet transfer. GTP-C is also used to transfer MM signalling messages between SGSNs, which corresponds to the MAP/G interface on the circuit switched side.

GTP-C operates on two different interfaces, Gn and Gp. The Gn interface is shared between GSNs (i.e. SGSNs and GGSNs) belonging to the UMTS packet domain network. The Gp interface is used for UMTS network interworking between two UMTS packet domain networks. See Figure 5.4.

On both Gn and Gp interfaces GTP-C operates on top of the UDP/IP protocol family. UDP provides connectionless message transfer without the need to establish a connection. The main service of IP is message routing between source and destination network elements.

In SGSN the GTP-C protocol entity interworks with the GTP-U protocol entity at the Gn side and with the GMM and SM protocol entities also. In GGSN the GTP-C entity interworks

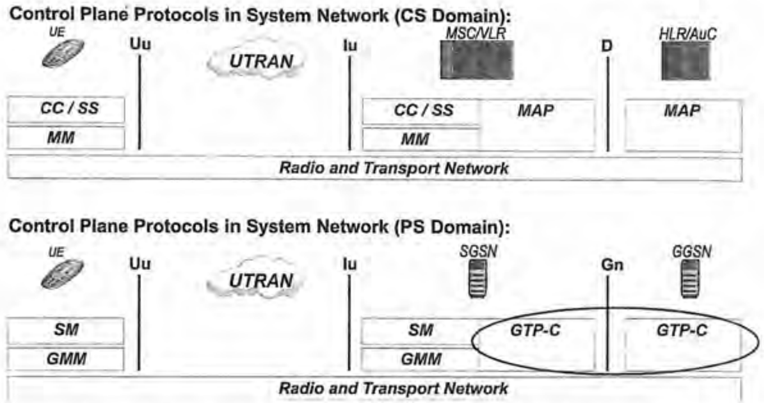


Figure 9.42. Control plane – GPRS tunnelling protocol for the control plane

with the GTP-U protocol entity located at the Gn interface and with the external network interworking function at the Gi interface, e.g. in order to allocate dynamic addresses for PDP contexts.

The GTP-C protocol is required to be able to transfer its signalling messages reliably over the Gn and Gp interfaces. Therefore the design of the GTP-C had to solve the problem of unreliable message transfer provided by the UDP/IP transport layers. Thus GTP-C includes timers and message sequence numbers needed to control and detect message loss and message retransmission.

GTP-C Procedures Between SGSN and GGSN

GTP-C is used to create, delete and modify GTP tunnels that are used to transfer user data packets. GTP-C also checks if the path towards the peer GTP-C is alive.

When a PDP context is activated, GTP-C in SGSN initiates GTP tunnel establishment towards GGSN. The purpose of the GTP tunnel establishment procedure is to negotiate parameters related to user data transfer between SGSN and GGSN. Such parameters are QoS and parameters associated with the activated address, like the traffic flow template for example.

When a PDP context is deactivated, GTP-C removes GTP tunnelling information from the SGSN and GGSN.

The PDP context modification procedure provides a flexible and dynamic way to react to changed conditions. When a PDP context is modified, e.g. due to changed QoS attributes or load condition, GTP-C initiates a GTP tunnel modification procedure. In the GTP tunnel modification procedure both the SGSN and GGSN are able to modify almost all the parameters negotiated in GTP tunnel establishment procedure. The PDP context modification procedure is also performed as part of inter-SGSN routing area update procedure and SRNS relocation procedure involving two SGSNs to notify GGSN of the new SGSN starting to serve the UE.

In the case of network requested PDP context activation is supported and GGSN does not include protocols needed to communicate with HLR across the Gc interface, GTP-C is used to transfer messages to another GSN that converts messages between GTP and MAP. MAP protocol is then used to communicate across the Gc interface towards HLR. The basic need for a GGSN to communicate with HLR is to find out if the UE is attached to the network and also the address of the SGSN serving the UE. This information is needed by the GGSN before it can request the SGSN to initiate a PDP context activation procedure.

GTP-C Procedures Between SGSNs

GTP-C is used in communication between two SGSNs in order to exchange information related to a UE. The communication is initiated when a UE makes a GPRS attach procedure, inter-SGSN routing area update procedure or UTRAN makes the decision to perform the SRNS relocation procedure involving two SGSNs.

The use of GTP-C in exchanging UE related information between SGSNs minimises the need to use a radio interface. For example when an inter-SGSN routing area update procedure or SRNS relocation procedure involving two SGSNs is performed, the new SGSN serving the UE receives almost all the needed information of the UE from the old SGSN. Such information consists of GMM related parameters, like IMSI and GMM context, and SM related parameters, like active PDP contexts.

The use of GTP-C in exchanging UE related information between SGSNs also strengthens the system security. For example when a UE makes a GPRS attach procedure, IMSI may be retrieved from the old SGSN and not from the UE over the radio interface.

The GTP-C protocol is defined in the 3GPP specification TS 29.060.

9.5.3. User Plane in the System Network

In the case of the CS domain the role of user plane protocols is taken by the speech codec

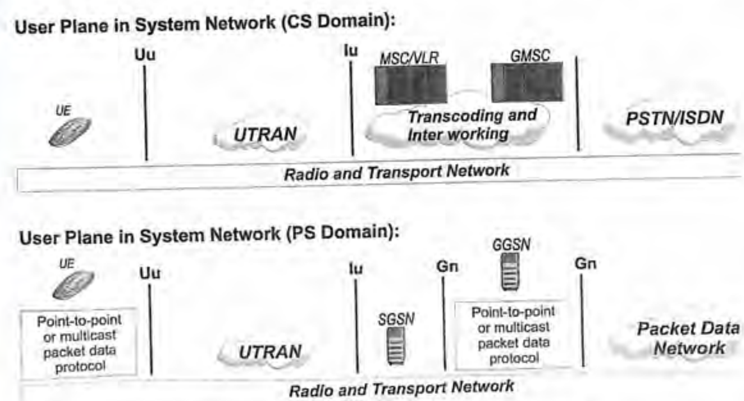


Figure 9.43. User plane in the system network

functions, which are active in the UE and serving MSC/VLR (Figure 9.43). In 3GPP R99 the predefined speech codec is Adaptive Multi Rate (AMR) codec, which was discussed in Chapter 7. Besides transcoding the interworking functions, mentioned already in Chapter 5, have to be taken care of between the UMTS and external PSTN/ISDN network.

In the case of the PS domain the UE and GGSN exchange data packets according to some point-to-point or multicast packed data protocol. For the time being the most common user plane protocol will continue to be the IP protocol.

10

Procedure Examples

In order to allow the reader to accomplish a network-wide view of some of the most important UMTS functionalities discussed in the preceding chapters we present a few examples of UMTS system procedures. These procedures illustrate the co-ordination of actions carried out by all network entities and the role of the UMTS protocols in controlling this co-ordination.

10.1. Elementary Procedures

In this chapter a basic model for system-wide procedures is presented by modelling each system procedure as a kind of communication-intensive transaction. The concept of *transaction* is used in order to underline the fact that these system-wide procedures are executed into well-defined completion and that subsequent procedures represent fairly independent instances of communication.

The layering and interworking of protocols as described in Chapter 9 makes it possible to distinguish a set of elementary procedures, which can be used as building blocks in the design of network transactions. Depending on a transaction and its type – whether it is mobile terminated or mobile originated – some parameters and messages vary inside the elementary procedures and some elementary procedures may or may not be used.

The UMTS protocol interworking model (see Figure 9.4) is needed here in the sense that different procedures make use of different layers in the UMTS network. The transport network is used in every procedure: signalling transport is always required and user data transport is required by many transactions. Radio network functions are needed whenever access network services are required within an elementary procedure. The overall control of system transactions is done by the system network protocols, which invoke the elementary procedures in a stepwise manner and determine how the transaction is proceeding through different steps.

Basically any network transaction can be divided into eight steps as presented in Figure 10.1. For each step an elementary procedure can be distinguished.

Paging is a mobility management procedure used when searching certain subscriber from the network coverage area. This procedure is only executed if the transaction originates from the network side. The remaining seven steps are the same whether the transaction is originated or terminated by the UE.

Radio Resource Control (RRC) connection set-up is an elementary procedure containing activities and message flow to establish a radio control connection between the terminal and radio access network. The details of this procedure vary depending on the case.