

**TSG-SA meeting #4
21 - 23 June 1999
Miami**

TSGS#4(99)322
Revised TSG#4(99)301

Title: 3G TR 23.907 version 1.1.0

Source: TSG SA WG2

Purpose: Information

3G TR 23.907 1.1.0 (1999-05)

Technical Report

3rd Generation Partnership Project; Technical Specification Group Services and System Aspects QoS Concept (3G TR 23.907 version 1.1.0)



The present document has been developed within the 3rd Generation Partnership Project (3GPP™) and may be further elaborated for the purposes of 3GPP.

The present document has not been subject to any approval process by the 3GPP Organisational Partners and shall not be implemented. This Specification is provided for future development work within 3GPP only. The Organisational Partners accept no liability for any use of this Specification.

Specifications and reports for implementation of the 3GPP™ system should be obtained via the 3GPP Organisational Partners' Publications Offices.

Reference

DTS/TSGS-0223101U

Keywords

<keyword[, keyword]>

3GPP

Postal address

3GPP support office address

650 Route des Lucioles - Sophia Antipolis
Valbonne - FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Internet

<http://www.3gpp.org>

Contents

Contents	3
Intellectual Property Rights	5
Foreword	5
Introduction.....	5
1 Scope	6
2 References	6
2.1 Normative References	6
2.2 Informative References.....	6
3 Definitions, symbols and abbreviations.....	7
3.1 Definitions	7
3.2 Abbreviations	7
4 High Level Requirements	8
4.1 End User QoS Requirements	8
4.2 General Requirements for QoS	8
4.3 Technical Requirements for QoS.....	9
5 QoS Capabilities of Radio Interface.....	9
6 QoS Architecture	10
6.1 Overview of Different Levels of QoS.....	10
6.1.1 The End-to-End Service and UMTS Bearer Service	10
6.1.2 The Radio Access Bearer Service and the Core Network Bearer Service.....	11
6.1.3 The Radio Bearer Service and the Iu Bearer Service.....	11
6.1.4 The Backbone Network Service	11
6.2 QoS Management Functions in the Network.....	12
6.2.1 Description of functions	12
6.2.1.1 QoS management functions for UMTS bearer service in the control plane	12
6.2.1.2 Functions for UMTS bearer service in the user plane	12
6.2.2 Allocation of QoS management functions	13
6.2.2.1 QoS management functions for UMTS bearer service in the control plane	13
6.2.2.2 QoS management functions for the UMTS bearer service in the user plane	14
6.3 UMTS QoS Classes.....	15
6.3.1 Conversational class.....	15
6.3.2 Streaming class	16
6.3.3 Interactive class	16
6.3.4 Background class	16
6.4 QoS Parameters	17
6.4.1 General	17
6.4.2 UMTS Bearer Service Attributes.....	17
6.4.2.1 List of attributes	17
6.4.2.2 Attributes discussed per class	19
6.4.2.3 UMTS bearer attributes: summary	20
6.4.3 Radio Access Bearer Service Attributes.....	21
6.4.3.1 List of attributes	21
6.4.3.2 Attributes discussed per class	22
6.4.3.3 Radio Access Bearer attributes: summary	23
6.4.4 Radio Bearer Service Parameters	23
6.4.5 Iu Bearer Service Parameters	23
6.4.6 Gn Bearer Service Parameters.....	23
7 QoS Parameter Mapping.....	23
7.1 Sources of UMTS Bearer Service Parameters.....	24

7.1	From Application Parameters to UMTS Bearer Service Parameters	24
7.2	From UMTS Bearer Service Parameters to Radio Access Bearer Service Parameters	24
7.3	From UMTS Bearer Service Parameters to CN Bearer Service Parameters	24
7.4	From Radio Access Bearer Service Parameters to Radio Bearer Service Parameters	24
7.5	From Radio Access Bearer Service Parameters to Iu Bearer Service Parameters	24
8	QoS in Connection Setup	25
9	Interworking.....	25
9.1	UMTS-GSM CS/GPRS	25
9.1.1	UMTS-GSM CS	25
9.1.2	UMTS-GPRS	25
9.2	UMTS-PSTN.....	26
9.3	UMTS-ISDN	26
9.4	UMTS-Internet	27
9.4.1	UMTS-Integrated Services	27
9.4.2	UMTS-Differentiated Services	27
	History	27

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in SR 000 314: "*Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards*", which is available **free of charge** from the ETSI Secretariat. Latest updates are available on the ETSI Web server (<http://www.etsi.fr/ipr> or <http://www.etsi.org/ipr>).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This draft Technical Report (TR) has been produced by the Special Mobile Group (SMG) of the European Telecommunications Standards Institute (ETSI).

This document is to define the basic issues related to the UMTS QoS concept.

The contents of this TR are subject to continuing work within SMG and may change following formal SMG approval. Should SMG modify the contents of this TR it will then be republished by ETSI with an identifying change of release date and an increase in version number as follows:

Version 1.x.y

where:

- 1 indicates sent to SMG for Information
- x the second digit is incremented for all other types of changes, i.e. technical enhancements, corrections, updates, etc.
- y the third digit is incremented when editorial only changes have been incorporated in the specification.

Introduction

1 Scope

Scope of this document is to provide the framework for Quality of Service in UMTS. The document shall be used as a living document which will cover all issues related Quality of Service in UMTS.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies.
- A non-specific reference to an ETS shall also be taken to refer to later versions published as an EN with the same number.

2.1 Normative References

- [1] UMTS 23.10 version 0.8.1, Universal Mobile Telecommunications System (UMTS), UMTS Access Stratum - Services and Functions
- [2] UMTS 22.25 version 3.1.0, Universal Mobile Telecommunication System (UMTS), Quality of Service and Network Performance
- [3] UMTS 22.01 version 3.2.1, Universal Mobile Telecommunications System (UMTS), Service aspects, Service principles
- [4] UMTS 22.60 version 3.0.0, Universal Mobile Telecommunications System (UMTS), Service aspects, Mobile multimedia services including mobile Intranet and Internet services
- [5] UMTS 21.06, Network and Service Management Requirements for UMTS
- [6] UMTS 23.20 version 0.7.0, Evolution of the GSM platform towards UMTS

2.2 Informative References

- [1] Working assumption on Radio Access Bearer Services, Version: 0.0.2, Date: 14 May 1998
- [2] ETSI SMG12 TDoc 98s721, Rome, Italy, 14 - 18 September, 1998, Source: Editor (Ericsson), Draft 0.0.7 of Requirements Relevant for UTRAN Architecture
- [3] ETSI SMG12 TDoc 98s767, Rome, Italy, 14 - 18 September, 1998, Source: Motorola, Change request No. A086 rev.1
- [4] ETSI SMG12 TDoc 98s754, Rome, Italy, 14 - 18 September, 1998, Source: Lucent Technologies, Proposal for ETE UMTS QoS Control Principles
- [5] ETSI SMG12 TDoc 98s790, Rome, Italy, 14 - 18 September, 1998, Source: Motorola, QoS Profile Parameters

- [6] ATM Traffic Management Specification Version 4.0 [TM 4.0]
[7] ATM User-Network Interface (UNI) Signalling Specification Version 4.0 [UNI 4.0]
[8] ETSI SMG12 Tdoc 98s789, Rome, Italy, 14-18 September, 1998, Source: Motorola, Enhancements to Radio Bearer Parameters

3 Definitions, symbols and abbreviations

3.1 Definitions

For the purpose of this document the following definitions apply.

3.2 Abbreviations

For the purpose of this document the following abbreviations apply.

3G	3rd Generation
AAL2	ATM Adaptation Layer 2
AAL5	ATM Adaptation Layer 5
ATM	Asynchronous Transfer Mode
CDMA	Code Division Multiple Access
CS	Circuit Switched
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communication
GTP	GPRS Tunneling Protocol
HO	Handover
IP	Internet Protocol
ISDN	Integrated Services Digital Network
L1	Layer 1, physical layer
L2	Layer 2, link layer
LLC	Logical Link Control
MAC	Medium Access Control
MO	Mobile Originating Call
MS	Mobile Station
MT	Mobile Terminating Call
nrt	non-real time
NSAP	Network layer Service Access Point
NT	Non-Transparent

PDP	Packet Data Protocol
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RAB	Radio Access Bearer
RAN	Radio Access Network
RLC	Radio Link Control
RSVP	Resource Reservation Protocol
rt	real time
SAP	Service Access Point
SDU	Service Data Unit
SNDCP	Sub-Network Dependent Convergence Protocol
UDP	User Datagram Protocol
T	Transparent
TCP	Transmission Control Protocol
TD-CDMA	Time Division CDMA
TE	Terminal Equipment
UMTS	Universal Mobile Telecommunication System
WCDMA	Wideband CDMA

4 High Level Requirements

4.1 End User QoS Requirements

Generally, end users care only the issues that are visible to them. The involvement of the user leads to the following conclusions. From the end-user point of view:

- Only the QoS perceived by end-user matter.
- The number of user defined/controlled parameters has to be as small as possible.
- Derivation/definition of QoS attributes from the application requirements has to be simple.
- QoS attributes must be able to support all applications that are used, a certain number of applications have the characteristic of asymmetric nature between two directions, uplink/downlink.
- QoS definitions have to be future proof.
- QoS has to be provided end-to-end.

4.2 General Requirements for QoS

- QoS parameters (or mapping of them) should not be restricted to one or few external QoS control

mechanisms but the QoS concept should be capable of providing different levels of QoS by using UMTS specific control mechanisms (not related to QoS mechanisms in the external networks).

- All parameters have to have unambiguous meaning.
- QoS mechanism have to allow efficient use of radio capacity.
- Allow independent evolution of Core and Access networks
- Allow evolution of UMTS network, (i.e., eliminate or minimise the impact of evolution of transport technologies in the wireline world)
- All parameter combinations have to have unambiguous meaning.

4.3 Technical Requirements for QoS

This chapter presents the general high-level technical requirements for the UMTS QoS. QoS will be defined with a set of parameters. These parameters should meet the following criteria:

- UMTS QoS control mechanisms shall provide QoS parameter control on a peer to peer basis between MS and 3G gateway node.
- The UMTS QoS mechanisms shall provide a mapping between application requirements and UMTS services.
- The UMTS QoS control mechanisms shall be able to efficiently interwork with current QoS schemes. Further, the QoS concept should be capable of providing different levels of QoS by using UMTS specific control mechanisms (not related to QoS mechanisms in the external networks).
- A session based approach needs to be adopted for all packet mode communication within the 3G serving node with which UMTS QoS approach must be intimately linked, essential features are multiple QoS streams per address.
- The UMTS shall provide a finite set of QoS definitions.
- The overhead and additional complexity caused by the QoS scheme should be kept reasonably low, as well as the amount of state information transmitted and stored in the network.
- QoS shall support efficient resource utilisation.
- The QoS parameters are needed to support asymmetric bearers.
- Applications (or special software in MS or 3G gateway node) should be able to indicate QoS values for their data transmissions.
- QoS behaviour should be dynamic , i.e., it shall be possible to modify QoS parameters during an active session.
- Number of parameters should be kept reasonably low (increasing number of parameters, increase system complexity).

5 QoS Capabilities of Radio Interface

[Editor note: This chapter shall describe the L1 from the QoS point of view. The intent of the chapter is to provide background to compare e.g. RAB parameters against L1 i.e. is L1 capable of utilising developed QoS concept.]

6 QoS Architecture

6.1 Overview of Different Levels of QoS

Network Services are considered end-to-end, this means from a Terminal Equipment (TE) to another TE. A End-to-End Service may have a certain Quality of Service (QoS) which is provided for the user of a network service. It is the user that decides whether he is satisfied with the provided QoS or not.

To realise a certain network QoS a Bearer Service with clearly defined characteristics and functionality is to be set up from the source to the destination of a service.

A bearer service includes all aspects to enable the provision of a contracted QoS. These aspects are among others the control signalling, user plane transport and QoS management functionality. A UMTS bearer service layered architecture is depicted in Figure 1, each bearer service on a specific layer offers it's individual services using services provided by the layers below.

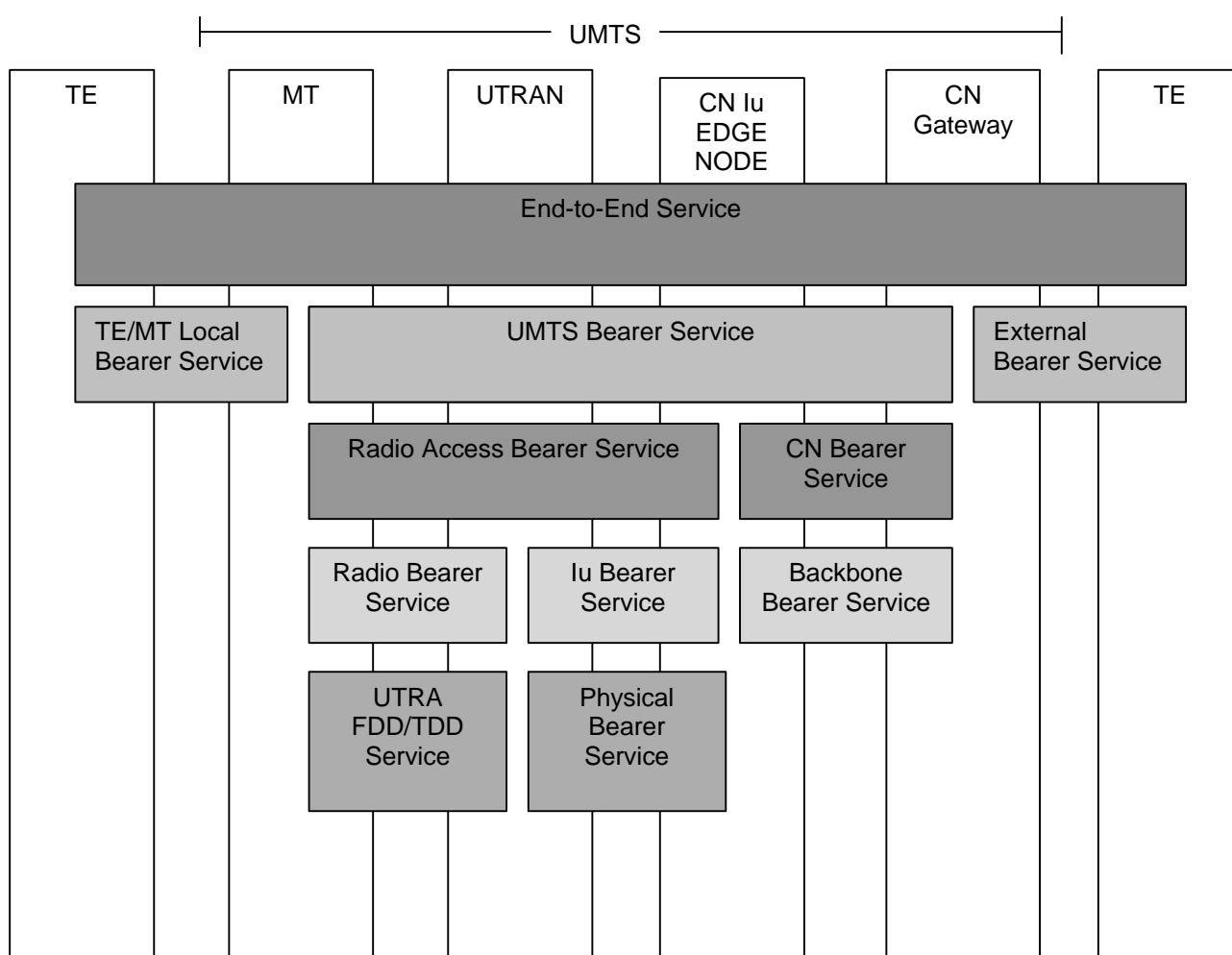


Figure 1: UMTS QoS Architecture

6.1.1 The End-to-End Service and UMTS Bearer Service

On its way from the TE to another TE the traffic has to pass different bearer services of the network(s). A TE is connected to the UMTS network by use of a Mobile Termination (MT). The *End-to-End Service* on the

application level uses the bearer services of the underlying network(s). As the *End-to-End Service* is conveyed over several networks (not only UMTS) it is not subject for further elaboration in this document.

The *End-to-End-Service* used by the TE will be realised using a *TE/MT Local Bearer Service*, a *UMTS Bearer Service*, and an *External Bearer Service*.

TE/MT Local Bearer Service is not further elaborated here as this bearer service is outside the scope of the UMTS network.

Having said that the *End-to-End Bearer Service* is beyond the scope of this document it is however the various services offered by the *UMTS Bearer Service* that the UMTS operator offers. It is this bearer service that provides the UMTS QoS.

The *External Bearer Service* is not further elaborated here as this bearer may be using several network services, e.g. another *UMTS Bearer Service*.

6.1.2 The Radio Access Bearer Service and the Core Network Bearer Service

As described in the previous chapter it is the UMTS Bearer Service that provides the UMTS QoS. The UMTS Bearer Service consists of two parts, the Radio Access Bearer Service and the Core Network Bearer Service. Both services reflect the optimised way to realise the UMTS Bearer Service over the respective cellular network topology taking into account such aspects as e.g. mobility and mobile subscriber profiles.

The Radio Access Bearer Service provides confidential transport of signalling and user data between MT and CN Iu Edge Node with the QoS adequate to the negotiated UMTS Bearer Service or with the default QoS for signalling. This service is based on the characteristics of the radio interface and is maintained for a moving MT.

The Core Network Bearer Service of the UMTS core network connects the UMTS CN Iu Edge Node with the CN Gateway to the external network. The role of this service is to efficiently control and utilise the backbone network in order to provide the contracted UMTS bearer service.

6.1.3 The Radio Bearer Service and the Iu Bearer Service

The Radio Access Bearer Service is realised by a Radio Bearer Service and an Iu-Bearer Service.

The role of the Radio Bearer Service is to cover all the aspects of the radio interface transport. This bearer service uses the UTRA FDD/TDD, which is not elaborated further in this document.

The Iu-Bearer Service together with the Physical Bearer Service provides the transport between UTRAN and CN.

6.1.4 The Backbone Network Service

The Core Network Bearer Service uses a generic Backbone Network Service.

The Backbone Network Service covers the layer 1/Layer2 functionality and is selected according to operator's choice in order to fulfil the QoS requirements of the Core Network Bearer Service. The Backbone Network Service is not specific to UMTS but may reuse an existing standard.

6.2 QoS Management Functions in the Network

The purpose of this chapter is to give a comprehensive overview of functionality needed to establish, modify and maintain a UMTS Bearer Service with a specific QoS. The QoS management functions need not necessarily all to be standardised. Their allocation to the UMTS entities shall indicate the requirement for the specific entity to enforce the QoS commitments negotiated for the UMTS bearer service. The specific realisation of these functions is implementation dependent and has only to maintain the specified QoS characteristics. The QoS management functions of all UMTS entities together shall ensure the provision of the negotiated service between the access points of the UMTS bearer service. The end-to-end service is provided by translation/mapping with UMTS external services.

6.2.1 Description of functions

6.2.1.1 QoS management functions for UMTS bearer service in the control plane

Service Manager co-ordinates the functions of the control plane for establishing, modifying and maintaining the service it is responsible for. And, it provides all user plane QoS management functions with the relevant attributes. The service manager offers services to other instances, it signals with peer service managers and uses services provided by other instances. The service manager may perform an attribute translation to request lower layer services. Furthermore, it may interrogate other control functions to receive permission for service provision.

Translation function converts between the internal service primitives for UMTS bearer service control and the various protocols for service control of interfacing external networks. The translation includes the converting between UMTS bearer service attributes and QoS parameters of the external networks service control protocol (e.g. between IETF TSPEC and UMTS service attributes). The service manager may include a translation function to convert between its service attributes and the attributes of a lower layer service it is using.

Admission/Capability control maintains information about all available resources of a network entity and about all resources allocated to UMTS bearer services. It determines for each UMTS bearer service request or modification whether the required resources can be provided by this entity and it reserves these resources if allocated to the UMTS bearer service. The function checks also the capability of the network entity to provide the requested service, i.e. whether the specific service is implemented and not blocked for administrative reasons. The resource control performed by the admission control supports also the service retention.

Subscription Control checks the administrative rights of the UMTS bearer service user to use the requested service with the specified QoS attributes.

6.2.1.2 Functions for UMTS bearer service in the user plane

User plane QoS management functions maintain the signalling and user data traffic within certain limits, defined by specific QoS attributes. UMTS bearer services with different QoS attribute values shall be supported by the QoS management functions. These functions ensure the provision of the QoS negotiated for a UMTS bearer service.

Mapping function provides each data unit with the specific marking required to receive the intended QoS at the transfer by a bearer service.

Classification function assigns data units to the established services of a MT according to the related QoS attributes if the MT has multiple UMTS bearer services established. The appropriate UMTS bearer service is derived from the data unit header or from traffic characteristics of the data.

Resource Manager distributes the available resources between all services sharing the same resource. The resource manager distributes the resources according to the required QoS. Example means for resource management are scheduling, bandwidth management and power control for the radio bearer.

Traffic conditioner provides conformance between the negotiated QoS for a service and the data unit traffic. Traffic conditioning is performed by policing or by traffic shaping. The policing function compares the data unit traffic with the related QoS attributes. Data units not matching the relevant attributes will be dropped or marked as not matching, for preferential dropping in case of congestion. The traffic shaper forms the data unit traffic according to the QoS of the service.

6.2.2 Allocation of QoS management functions

6.2.2.1 QoS management functions for UMTS bearer service in the control plane

The QoS management functions for controlling the UMTS bearer service are shown in figure 2. These control functions support the establishment and the modification of a UMTS bearer service by signalling/negotiation with the UMTS external services and by the establishment or modification of all UMTS internal services with the required characteristics.

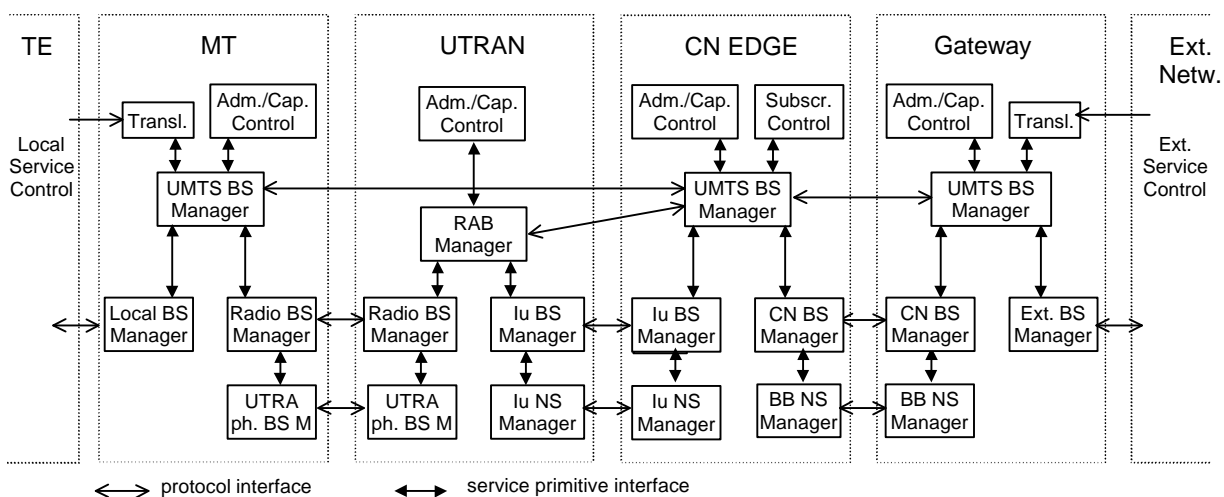


Figure 2: QoS management functions for UMTS bearer service in the control plane

Editors note: all logical interfaces are shown in the figure, contributions are expected to describe all interfaces; some interfaces may only apply to one (CS / PS) domain

The translation functions (Trans.) in the MT and the Gateway convert between external service signalling and internal service primitives including the translation of the service attributes. The translation function in the Gateway is FFS regarding packet oriented services.

The UMTS BS manager in the MT, CN EDGE and the Gateway signal between each other and via the translation function with external instances to establish or modify a UMTS bearer service. Each of the UMTS BS managers interrogates its associated admission/capability control whether the network entity supports the specific requested service and whether the required resources are available. Additionally, the CN EDGE UMTS BS manager verifies with the subscription control the administrative rights for using the service.

The UMTS BS manager of the MT translates the UMTS bearer service attributes into attributes for the local bearer service and requests this service from the local BS manager.

The UMTS BS manager of the CN EDGE translates the UMTS bearer service attributes into RAB service

attributes and lu bearer service attributes and it translates UMTS bearer service attributes into CN bearer service attributes. Also, the UMTS BS manager of the CN EDGE requests its lu BS manager, its CN BS manager and the RAB manager in the UTRAN to provide the required services.

The RAB manager verifies with its admission/capability control whether the UTRAN supports the specific requested service and whether the required resources are available. It translates the RAB service attributes into radio bearer service and lu bearer service attributes and requests the radio BS manager and the lu BS manager to provide bearer services with the required attributes.

The Gateway UMTS BS manager translates the UMTS bearer service attributes into CN bearer service attributes and requests its CN BS manager to provide the service. Furthermore, it translates the UMTS bearer service attributes into the external bearer service attributes and requests this service from the external BS manager.

Radio, lu and CN BS managers use services provided by lower layers as indicated in figure 2.

6.2.2.2 QoS management functions for the UMTS bearer service in the user plane

The QoS management functions of the UMTS BS for the user plane are shown in figure 3. These functions maintain the data transfer characteristics according to the commitments established by the UMTS BS control functions and expressed by the bearer service attributes. The QoS management user plane functions are provided with the relevant attributes by the QoS management control functions.

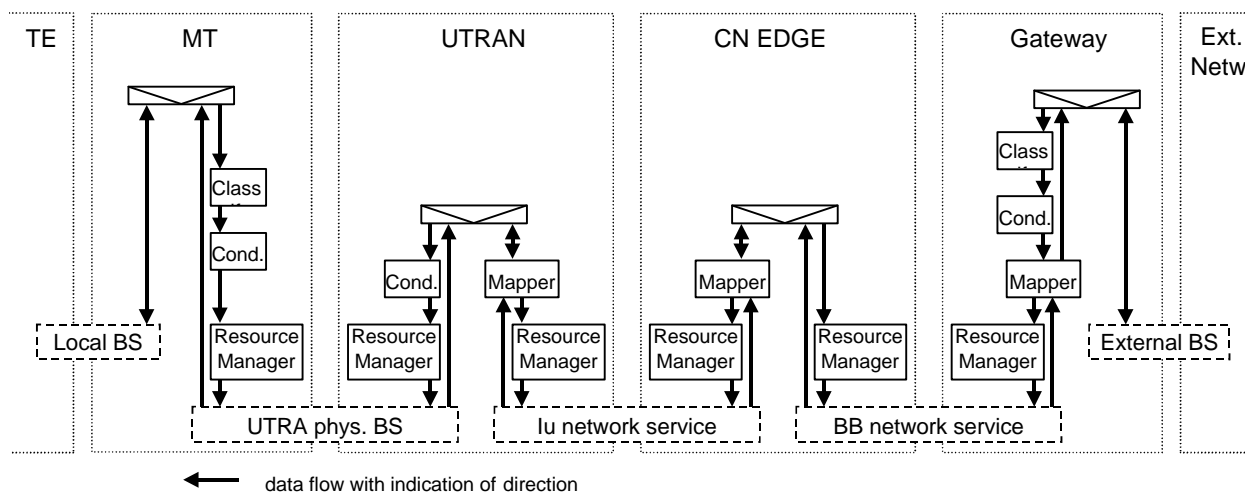


Figure 3: QoS management functions for the UMTS bearer service in the user plane

The classification function (Class.) in the Gateway and in the MT assign user data units received from the external bearer service or the local bearer service to the appropriate UMTS bearer service according to the QoS requirements of each user data unit. The classification function in the MT is FFS.

The traffic conditioner (Cond.) in the MT provides conformance of the uplink user data traffic with the QoS attributes of the relevant UMTS bearer service. In the Gateway a traffic conditioner may provide conformance of the downlink user data traffic with the QoS attributes of the relevant UMTS bearer service. The packet oriented transport of the downlink data units from the external bearer service to the UTRAN and the buffering in the UTRAN may result in bursts of downlink data units not conformant with the UMTS BS QoS attributes. A traffic conditioner in the UTRAN forms this downlink data unit traffic according to the relevant QoS attributes.

Editors note: Contributions are invited for the discussion whether the network can trust MTs or uplink conditioning in the UTRAN is required to prevent MTs from using more resources than allowed; physical bearer already limits MT's capabilities in misbehaving

The traffic conditioners are not necessarily separated functions. For example a resource manager may also provide conformance with the relevant QoS attributes by appropriate data unit scheduling. Or, if fixed resources are dedicated to one bearer service the resource limitations implicitly condition the traffic.

The mapping function marks each data unit with the specific QoS indication related to the bearer service performing the transfer of the data unit.

Each of the resource managers of a network entity is responsible for a specific resource. The resource manager distributes its resources between all bearer services requesting transfer of data units on these resources. Thereby, the resource manager attempts to provide the QoS attributes required for each individual bearer service.

6.3 UMTS QoS Classes

When defining the UMTS QoS classes the restrictions and limitations of the air interface have to be taken into account. It is not reasonable to define complex mechanisms as have been in fixed networks due to different error characteristics of the air interface. The QoS mechanisms provided in the cellular network have to be robust and capable of providing reasonable QoS resolution. Table 1 illustrates proposed QoS classes for UMTS.

In the proposal there are four different QoS classes (or traffic classes):

- Conversational class,
- Streaming class,
- Interactive class and
- Background class.

The main distinguishing factor between these classes is how delay sensitive the traffic is: Conversational class is meant for traffic which is very delay sensitive while Background class is the most delay insensitive traffic class.

Conversational and Streaming classes are mainly intended to be used to carry real-time traffic flows. The main divider between them is how delay sensitive the traffic is. Conversational real-time services, like video telephony, are the most delay sensitive applications and those data streams should be carried in Conversational class.

Interactive class and Background are mainly meant to be used by traditional Internet applications like WWW, Email, Telnet, FTP and News. Due to looser delay requirements, compare to conversational and streaming classes, both provide better error rate by means of channel coding and retransmission. The main difference between Interactive and Background class is that Interactive class is mainly used by interactive applications, e.g. interactive Email or interactive Web browsing, while Background class is meant for background traffic, e.g. background download of Emails or background file downloading. Responsiveness of the interactive applications is ensured by separating interactive and background applications. Traffic in the Interactive class has higher priority in scheduling than Background class traffic, so background applications use transmission resources only when interactive applications do not need them. This is very important in wireless environment where the bandwidth is low compared to fixed networks.

6.3.1 Conversational class

The most well known use of this scheme is telephony speech (e.g. GSM). But with Internet and multimedia a number of new applications will require this scheme, for example voice over IP and video conferencing tools. Real time conversation is always performed between peers (or groups) of live (human) end-users. This is the only scheme where the required characteristics are strictly given by human perception.

Real time conversation scheme is characterised by that the transfer time must be low because of the conversational nature of the scheme and at the same time that the time relation (variation) between information entities of the stream must be preserved in the same way as for real time streams. The maximum transfer delay is given by the human perception of video and audio conversation. Therefore the limit for acceptable transfer delay is very strict, as failure to provide low enough transfer delay will result in unacceptable lack of quality. The transfer delay requirement is therefore both significantly lower and more stringent than the round trip delay of the interactive traffic case.

Real time conversation - fundamental characteristics for QoS:

- preserve time relation (variation) between information entities of the stream
- conversational pattern (stringent and low delay)

6.3.2 Streaming class

When the user is looking at (listening to) real time video (audio) the scheme of real time streams applies. The real time data flow is always aiming at a live (human) destination. It is a one way transport.

This scheme is one of the newcomers in data communication, raising a number of new requirements in both telecommunication and data communication systems. It is characterised by that the time relations (variation) between information entities (i.e. samples, packets) within a flow must be preserved, although it does not have any requirements on low transfer delay.

The delay variation of the end-to-end flow must be limited, to preserve the time relation (variation) between information entities of the stream. But as the stream normally is time aligned at the receiving end (in the user equipment), the highest acceptable delay variation over the transmission media is given by the capability of the time alignment function of the application. Acceptable delay variation is thus much greater than the delay variation given by the limits of human perception.

Real time streams - fundamental characteristics for QoS:

- preserve time relation (variation) between information entities of the stream

6.3.3 Interactive class

When the end-user, that is either a machine or a human, is on line requesting data from remote equipment (e.g. a server), this scheme applies. Examples of human interaction with the remote equipment are: web browsing, data base retrieval, server access. Examples of machines interaction with remote equipment are: polling for measurement records and automatic data base enquiries (tele-machines).

Interactive traffic is the other classical data communication scheme that on an overall level is characterised by the request response pattern of the end-user. At the message destination there is an entity expecting the message (response) within a certain time. Round trip delay time is therefore one of the key attributes. Another characteristic is that the content of the packets must be transparently transferred (with low bit error rate).

Interactive traffic - fundamental characteristics for QoS:

- request response pattern
- preserve payload content

6.3.4 Background class

When the end-user, that typically is a computer, sends and receives data-files in the background, this scheme applies. Examples are background delivery of E-mails, SMS, download of databases and reception of measurement records.

Background traffic is one of the classical data communication schemes that on an overall level is characterised by that the destination is not expecting the data within a certain time. The scheme is thus more or less delivery time insensitive. Another characteristic is that the content of the packets must be transparently transferred (with low bit error rate).

Background traffic - fundamental characteristics for QoS:

- the destination is not expecting the data within a certain time
- preserve payload content

Table 1: UMTS QoS classes

Traffic class	Conversational class conversational RT	Streaming class streaming RT	Interactive class Interactive best effort	Background Background best effort
Fundamental characteristics	<ul style="list-style-type: none"> • Preserve time relation (variation) between information entities of the stream • Conversational pattern (stringent and low delay) 	<ul style="list-style-type: none"> • Preserve time relation (variation) between information entities of the stream 	<ul style="list-style-type: none"> • Request response pattern • Preserve payload content 	<ul style="list-style-type: none"> • Destination is not expecting the data within a certain time • Preserve payload content
Example of the application	- voice	- streaming video	- Web browsing	- background download of emails

6.4 QoS Parameters

6.4.1 General

The parameters related to throughput/bitrate should be separated for uplink/downlink in order to support asymmetric bearers.

6.4.2 UMTS Bearer Service Attributes

6.4.2.1 List of attributes

[Note: The text within square brackets explaining the purpose of each attribute can be excluded later if that information is given elsewhere in the technical report.]

Potential purpose and definition of a possible additional attribute 'average bitrate' is FFS.

Traffic class ['conversational', 'streaming', 'interactive', 'background']

Definition: type of application for which the UMTS bearer service is optimised

[Purpose: By including the traffic class itself as an attribute, UMTS can make assumptions about the traffic source and optimise the transport for that traffic type.]

Maximum bitrate [kbps]

Definition: maximum number of bits delivered by UMTS at a SAP within a measurement period, divided by the duration of the measurement period. The definition of the period is FFS.

[Purpose: Maximum bitrate can be used to make code reservations in the downlink of the radio interface. Its purpose is to limit the delivered bitrate to applications or external networks with such limitations]

Guaranteed bitrate [kbps]

Definition: guaranteed number of bits delivered by UMTS at a SAP within a measurement period (provided that there is data to deliver), divided by the duration of the measurement period. The definition of the measurement period is FFS.

[Purpose: Guaranteed bitrate may be used to facilitate admission control based on available resources, and for resource allocation within UMTS. Quality requirements expressed by e.g. delay and reliability attributes only apply to incoming traffic up to the guaranteed bitrate.]

Delivery order [y/n]

Definition: indicates whether the UMTS bearer shall provide in-sequence SDU delivery or not.

[Purpose: the attribute is derived from the user protocol [PDP type] and specifies if out-of-sequence SDUs are acceptable or not. This information cannot be extracted from the traffic class. Whether out-of-sequence SDUs are dropped or re-ordered depends on the specified reliability]

SDU size information [bits]

Definition: list of possible exact sizes of SDUs

[Purpose: UTRAN needs SDU size information to be able to operate in transparent RLC protocol mode, which is beneficial to spectral efficiency and delay when RLC re-transmission is not used. Thus, if the application can specify SDU sizes, the bearer is less expensive.]

Reliability

Definition: information on the reliability in terms of error ratio etc that the UMTS bearer provides. The error ratio is the relation between erroneous SDUs/bits and SDUs/bits requested to be transferred. Whether one or both of SDU and bit error ratio apply to a certain UMTS bearer is FFS. Additionally, reliability specifies whether the UMTS bearer shall deliver SDUs with detected errors or discard these SDUs. Whether this is implicitly given by the two above mentioned error ratios is FFS.

[Purpose: used to specify the reliability needed by the application. Especially UTRAN may set internal parameters based on this value; the reliability impact on the CN is assumed to be negligible.]

Transfer delay [s]

Definition: time between request to transfer an SDU at one SAP to its delivery at the other SAP. Transfer delay is specified for one or more fixed SDU sizes. Exact statistical transfer delay definition and fixed SDU sizes are FFS.

[Purpose: used to specify the delay tolerated by the application. It allows UTRAN to set transport formats and

ARQ parameters.]

[Note: Transfer delay of an arbitrary SDU is not meaningful for a bursty source, since the last SDUs of a burst may have long delay due to queuing, whereas the meaningful response delay perceived by the user is the delay of the first SDU of the burst.]

Traffic handling priority

Definition: specifies the relative importance for handling of all SDUs belonging to the UMTS bearer compared to the SDUs of other bearers.

[Purpose: Within the interactive class, there is a definite need to differentiate between bearer qualities. This is handled by using the traffic handling priority attribute, to allow UMTS to schedule traffic accordingly. By definition, priority is an alternative to absolute guarantees, and thus these two attribute types cannot be used together for a single bearer.]

Allocation/Retention Priority

Definition: specifies the relative importance compared to other UMTS bearers for allocation and retention of the UMTS bearer.

[Purpose: Priority is used for differentiating between bearers when performing allocation and retention of a bearer, and the value is typically related to the subscription.]

6.4.2.2 Attributes discussed per class

Conversational class

Although the bitrate of a conversational source codec may vary, conversational traffic is assumed to be relatively non-bursty. **Maximum bitrate** specifies the upper limit of the bitrate with which the UMTS bearer delivers SDUs at the SAPs. The UMTS bearer is not required to transfer traffic exceeding the **Guaranteed bitrate**. Maximum and guaranteed bitrate attributes are used for resource allocation within UMTS. Minimum resource requirement is determined by guaranteed bitrate (When a conversational source generates less traffic than allocated for the bearer, the unused resources can of course be used by other bearers.)

Since the traffic is non-bursty, it is meaningful to guarantee a **transfer delay** of an arbitrary SDU.

Conversational bearers are likely to be realised in UTRAN without RLC re-transmissions. Hence, UTRAN transport is more efficient and thereby cheaper if RLC PDU size is adapted to UMTS bearer SDU size (RLC transparent mode). This motivates the use of **SDU size information**.

By using the **reliability** attribute(s), the application requirement on error rate can be specified, as well as whether the application wants UMTS to detect and discard SDUs containing errors and an adequate forward error correction means can be selected.

Streaming class

As for conversational class, streaming traffic is assumed to be rather non-bursty. **Maximum bitrate** specifies the upper limit of the bitrate the UMTS bearer delivers SDUs at the SAPs. The UMTS bearer is not required to transfer traffic exceeding the Guaranteed bitrate. Maximum and guaranteed bitrate attributes are used for resource allocation within UMTS. Minimum resource requirement is determined by guaranteed bitrate. (When a streaming source generates less traffic than allocated for the bearer, the unused resources can of course be used by other bearers.)

Since the traffic is non-bursty, it is meaningful to guarantee a **transfer delay** of an arbitrary SDU.

The transfer delay requirements for streaming are typically in a range where at least in a part of this range

RLC re-transmission may be used. It is assumed that the application's requirement on delay variation is expressed through the transfer delay attribute, which implies that there is no need for an explicit delay variation attribute.

By using the **reliability** attribute(s), the application requirement on error rate can be specified, as well as whether the application wants UMTS to detect and discard SDUs containing errors.

Interactive class

This bearer class is optimised for transport of human or machine interaction with remote equipment, such as web browsing. The source characteristics are unknown but may be bursty.

To be able to limit the delivered data rate for applications and external networks by traffic conditioning, **maximum bitrate** is included.

There is a definite need to differentiate between quality for bearers within the interactive class. One alternative would be to set absolute guarantees on delay, bitrate etc, which however at present seems complex to implement within UTRAN/CN. Instead, **traffic handling priority** is used. SDUs of a UMTS bearer with higher traffic handling priority is given priority over SDUs of other bearers within the interactive class, through UMTS-internal scheduling.

It is principally impossible to combine this relative approach with attributes specifying delay, bitrate, packet loss etc, so an interactive bearer gives no quality guarantees, and the actual bearer quality will depend on the load of the system and the admission control policy of the network operator.

The only additional attribute that is reasonable to specify is the bit integrity of the delivered data, which is a part of the **reliability**.

Background class

The background class is optimised for machine-to-machine communication that is not delay sensitive, such as messaging services. Background applications tolerate a higher delay than applications using the interactive class, which is the main difference between the background and interactive classes.

UMTS only transfers background class SDUs when there is definite spare capacity in the network. To be able to limit the delivered data rate for applications and external networks by traffic conditioning, **maximum bitrate** is included.

No other guarantee than bit integrity in the delivered data, according to the **reliability**, is needed.

6.4.2.3 UMTS bearer attributes: summary

In Table 2, the defined UMTS bearer attributes and their relevancy for each bearer class are summarised. Observe that traffic class is an attribute itself.

Table 2. UMTS bearer attributes defined for each bearer class.

Traffic class	Conversational class	Streaming class	Interactive class	Background class
Maximum bitrate	X	X	X	X
Delivery order	X	X	X	X
SDU size information	X			
Reliability	X	X	X	X

Transfer delay	X	X		
Guaranteed bit rate	X	X		
Traffic handling priority			X	
Allocation/Retention priority	X	X	X	X

6.4.3 Radio Access Bearer Service Attributes

6.4.3.1 List of attributes

[Note: The text within square brackets explaining the purpose of each attribute can be excluded later if that information is given elsewhere in the technical report.]

Traffic class ['conversational', 'streaming', 'interactive', 'background']

Definition: type of application for which the Radio Access Bearer service is optimized

[Purpose: By including the traffic class itself as an attribute, UTRAN can make assumptions about the traffic source and optimize the transport for that traffic type. In particular, buffer allocation may be based on traffic class.]

Maximum bitrate [kbps]

Definition: maximum number of bits delivered by UTRAN at a SAP within a measurement period, divided by the duration of the measurement period. The definition of the period is FFS.

[Purpose: to limit the delivered bitrate to applications or external networks with such limitations]

Guaranteed bitrate [kbps]

Definition: guaranteed number of bits delivered at a SAP within a measurement period (provided that there is data to deliver), divided by the duration of the measurement period. The definition of the measurement period is FFS.

[Purpose: Guaranteed bitrate may be used to facilitate admission control based on available resources, and for resource allocation within UTRAN. Quality requirements expressed by e.g. delay and reliability attributes only apply to incoming traffic up to the guaranteed bitrate. The guaranteed bitrate at the RAB level may be different from that on UMTS bearer level, for example due to header compression.]

Delivery order [y/n]

Definition: indicates whether the UMTS bearer shall provide in-sequence SDU delivery or not.

[Purpose: specifies if out-of-sequence SDUs are acceptable or not. This information cannot be extracted from the traffic class. Whether out-of-sequence SDUs are dropped or re-ordered depends on the specified reliability]

SDU size information [bits]

Definition: list of possible exact sizes of SDUs.

[Purpose: UTRAN needs SDU size information to be able to operate in transparent RLC protocol mode, which is beneficial to spectral efficiency and delay when RLC re-transmission is not used. Thus, if the application can specify SDU sizes, the bearer is less expensive.]

Reliability

Definition: information on the reliability in terms of error ratio etc that the Radio Access Bearer provides. The error ratio is the relation between erroneous SDUs/bits and SDUs/bits requested to be transferred. Whether one or both of SDU and bit error ratio apply to a certain Radio Access bearer is FFS. Additionally, reliability specifies whether the Radio Access bearer shall deliver SDUs with detected errors or discard these SDUs. Whether this is implicitly given by the two above mentioned error ratios is FFS.

[Purpose: UTRAN may set internal parameters based on specified reliability.]

Transfer delay [s]

Definition: time between request to transfer an SDU at one SAP to its delivery at the other SAP. Transfer delay is specified for one or more fixed SDU sizes. Exact statistical transfer delay definition and fixed SDU sizes are FFS.

[Purpose: specifies the UTRAN part of the total transfer delay for the UMTS bearer. It allows UTRAN to set transport formats and ARQ parameters.]

Traffic handling priority

Definition: specifies the relative importance for handling of all SDUs belonging to the radio access bearer compared to the SDUs of other bearers.

[Purpose: Within the interactive class, there is a definite need to differentiate between bearer qualities. This is handled by using the traffic handling priority attribute, to allow UTRAN to schedule traffic accordingly. By definition, priority is an alternative to absolute guarantees, and thus these two attribute types cannot be used together for a single bearer.]

Allocation/Retention Priority

Definition: specifies the relative importance compared to other Radio access bearers for allocation and retention of the Radio access bearer.

[Purpose: Priority is used for differentiating between bearers when performing allocation and retention of a bearer, and the value is typically related to the subscription.]

6.4.3.2 Attributes discussed per class

[Note: the use of new attributes introduced on Radio Access Bearer service level will be described here.]

Conversational class

Streaming class

Interactive class

Background class

6.4.3.3 Radio Access Bearer attributes: summary

In Table 3, the defined Radio Access Bearer attributes and their relevancy for each bearer class are summarized. Observe that traffic class is an attribute itself.

Table 3. Radio Access Bearer attributes defined for each bearer class.

Traffic class	Conversational class	Streaming class	Interactive class	Background class
Maximum bitrate	X	X	X	X
Delivery order	X	X	X	X
SDU size information	X			
Reliability	X	X	X	X
Transfer delay	X	X		
Guaranteed bit rate	X	X		
Traffic handling priority			X	
Allocation/ Retention priority	X	X	X	X

6.4.4 Radio Bearer Service Parameters

6.4.5 Iu Bearer Service Parameters

6.4.6 Gn Bearer Service Parameters

7 QoS Parameter Mapping

[Editor note: This chapter shall contain information of parameter mapping i.e. how parameter in different levels of QoS (from external world and within UMTS network) shall be mapped. Current sub-chapter division is based on an assumption that levels of QoS presented in chapter 6, but they are open to discussions.]

7.1 Sources of UMTS Bearer Service Parameters

UMTS bearer service parameters describe the service provided by the UMTS network to the user of the UMTS bearer service. A set of QoS parameters (QoS profile) specifies this service. At UMTS bearer service establishment or modification different QoS profiles have to be taken into account:

- The UE capabilities form a QoS profile which may limit the UMTS bearer service which can be provided.
- The UE or the terminal equipment (TE) within the terminating network may request a QoS profile at UMTS bearer establishment or modification. The application using the UE may request the UE to provide a UMTS bearer service with a specific QoS profile. If the application requests no specific QoS the UE may use a QoS profile configured within the UE (e.g., by AT commands). How the TE derives a QoS profile is out of scope for UMTS.
- A QoS profile in the UMTS subscription describes the upper limits for the provided service if the service user requests specific values. Otherwise, this QoS profile may describe a default QoS service profile requested by the user.
- Default QoS profile(s) may be configured by the operator for the UMTS bearer services provided by the network.
- A Network specific QoS profile characterising for example the current resource availability or other network capabilities or limitations may limit the provided UMTS bearer service or initiate a modification of an established UMTS bearer service.

7.1 From Application Parameters to UMTS Bearer Service Parameters

7.2 From UMTS Bearer Service Parameters to Radio Access Bearer Service Parameters

7.3 From UMTS Bearer Service Parameters to CN Bearer Service Parameters

7.4 From Radio Access Bearer Service Parameters to Radio Bearer Service Parameters

7.5 From Radio Access Bearer Service Parameters to Iu Bearer Service Parameters

8 QoS in Connection Setup

[Editor note: This chapter shall contain information of QoS parameters when setting up a PDP context, bearer, etc.]

9 Interworking

The model for the UMTS QoS classes and parameters may not be any existing network or QoS protocol/mechanisms as such. The main goal of the specification is not to copy existing QoS mechanisms but rather to create a future proof concept that will provide means to transport different types of data with different QoS requirements. Thus the interworking of UMTS and existing network technologies has to be ensured. This chapter presents the most common technologies that UMTS shall be capable to interwork with.

9.1 UMTS-GSM CS/GPRS

The development of the GSM/GPRS QoS classes and parameters is going on.

9.1.1 UMTS-GSM CS

The mapping between UMTS-GSM CS is easy because in GSM CS services only channel type (NT/T) and bandwidth (number of used time slots) can vary. During handover between UMTS-GSM only reliability, delay and bandwidth are meaningful parameters.

HO from UMTS to GSM:

Reliability and delay together determines if NT or T service is used. Bandwidth is mapped to the next higher possible step.

HO from GSM to UMTS:

Type of service determines the required reliability. Bandwidth is mapped to the next higher possible step.

9.1.2 UMTS-GPRS

[Editor note: Part of GPRS phase 1 QoS are vaguely defined. This chapter has to be updated according to CRs to GPRS phase 1 QoS parameters. Definition of GPRS phase 2 is starting and it has to be taken into account in here and vice versa.]

GPRS has more QoS parameters than GSM CS thus requiring more complex mapping rules. Below an example of mapping GPRS phase 1 QoS parameters to UMTS traffic classes is presented.

Conversational Class

Conversational class services are mainly for conversational real time use. An example of conversational real time application is video telephony.

An appropriate use of GPRS parameters:

- Mean Throughput Class = Peak bit rate (constant bit rate)

Or Mean Throughput Class < Peak bit rate (variable bit rate)

- Reliability: 4 or 5 (no retransmissions)
- Precedence: 1-3
- Delay Class: 1 (real time)

Streaming Class

Streaming class services are mainly appropriate for streaming real time applications, e.g. video downloading. Some variation in delay can be tolerated because of application level buffering.

An appropriate use of GPRS parameters:

- Mean Throughput Class = Peak bit rate (constant bit rate)

Or Mean Throughput Class < Peak bit rate (variable bit rate)

- Reliability: 3 (light retransmissions)
- Precedence: 1-3
- Delay Class: 1 (real time)

Interactive Class

Interactive class services are mainly for interactive services requiring a variable guaranteed throughput: specialised applications (banking, plane reservation, ...), interactive WWW, Telnet etc.

An appropriate use of GPRS parameters:

- Mean Throughput Class has no meaning in UMTS
- Reliability: 1-2
- Precedence: 1-3
- Delay class: 2-4

Background Class

Background services are mainly for best effort services: background download, emails, calendar, event etc.

An appropriate use of GPRS parameters:

- Mean Throughput Class has no meaning in UMTS
- Reliability: 2
- Precedence: 1-3
- Delay class: 4 (best effort)

9.2 UMTS-PSTN

PSTN does not have QoS mechanisms thus QoS parameter interworking/mapping is not needed. However, means for determining required bandwidth, delay and reliability has to be developed. It is simple in MO cases but in MT cases the mechanisms (or in worst case defaults) have to be developed.

9.3 UMTS-ISDN

ISDN does not have QoS mechanisms thus QoS parameter interworking/mapping is not needed. However, means for determining required bandwidth, delay and reliability has to be developed. It is simple in MO cases

but in MT cases the mechanisms (or in worst case defaults) have to be developed.

9.4 UMTS-Internet

In the case of Internet applications, the selection of the class and appropriate traffic attribute values is made according to the Internet QoS parameters. Internet applications do not directly use the services of UMTS but they use Internet QoS definitions and attributes, which are mapped to UMTS QoS attributes at API. Currently there are two main Internet QoS concepts, namely Integrated Services and Differentiated Services. The mapping between Internet QoS and UMTS QoS is presented in following chapters.

IP based QoS models must be supported for PDP contexts, meaning both Integrated Services (IntServ) signalled by RSVP [RFC2205] and Differentiated Services (6-bit QoS parameter on each IP packet, DiffServ). Both mechanisms are controlled by applications residing in the TE, allowing different application specific QoS levels for the same PDP context. Application level IP based QoS must be mapped to UMTS packet core QoS by a network element at the border of the network, such as the 3G gateway node. RSVP support would require flow establishment, and possibly aggregation of flows, within the UMTS packet core network. Differentiated services would require that there is either one QoS profile for each traffic type or alternatively the priority and traffic type information is included in the data packets.

9.4.1 UMTS-Integrated Services

RSVP is an end-to-end control protocol for Integrated Services Architecture and requires (soft) connection state maintenance in each network element along the data path.

9.4.2 UMTS-Differentiated Services

Work in IETF focuses on Differentiated services where the plan is to specify the details of IP header bit usage for QoS differentiated classes of service. Proposals in the IETF include varying number of pre-defined bits for delay and dropping priorities allowing each packet to be handled independently based on the header bits. The main objective is to specify only a QoS mechanism into header fields rather than end-to-end flow management.

History

Document History		
October 1998		Scope agreed
November 1998	0.1.0	ToR agreed
January 1999	0.2.0	Parameters, traffic classes and management functions added
April 1999	0.3.0	Moved to template
April 1999	0.4.0	Parameter definitions and sources added and annexes removed
May 1999	0.5.0	RAB attributes added

May 1999	1.0.0	Version approved to be changed to 1.0.0 by TSG SA2 #5
June 1999	1.1.0	QoS management functions updated