ETSI TS 123 107 V10.1.0 (2011-06)

Technical Specification

Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); LTE; Quality of Service (QoS) concept and architecture (3GPP TS 23.107 version 10.1.0 Release 10)

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00   Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

Individual copies of the present document can be downloaded from:
http://www.etsi.org

The present document may be made available in more than one electronic version or in print. In any case of existing or perceived difference in contents between such versions, the reference version is the Portable Document Format (PDF). In case of dispute, the reference shall be the printing on ETSI printers of the PDF version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
http://portal.etsi.org/tb/status/status.asp

If you find errors in the present document, please send your comment to one of the following services:
http://portal.etsi.org/chaircor/ETSI_support.asp

Copyright Notification

ETSI

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 ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://webapp.etsi.org/IPR/home.asp).

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 ETSI 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 Technical Specification (TS) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

The present document may refer to technical specifications or reports using their 3GPP identities, UMTS identities or GSM identities. These should be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between GSM, UMTS, 3GPP and ETSI identities can be found under http://webapp.etsi.org/key/queryform.asp.

Contents

Foreword

This Technical Specification (TS) has been produced by the 3rd Generation Partnership Project (3GPP).

The present document identifies the Quality of Service (QoS) aspects for the 3GPP system.

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x   the first digit:

1   presented to TSG for information;

2   presented to TSG for approval;

3   or greater indicates TSG approved document under change control.

y   the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z   the third digit is incremented when editorial only changes have been incorporated in the document.

1 Scope

The present document provides the framework for Quality of Service within the 3GPP system. The main purpose is to specify the list of attributes applicable to the UMTS Bearer Service and the Radio Access Bearer Service, as well as describe the Quality of Service architecture to be used in the 3GPP system.

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. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1]   3GPP TS 23.110: "UMTS Access Stratum - Services and Functions".

[2]   3GPP TS 22.100: "UMTS Phase 1".

[3]   3GPP TS 23.121: "Architectural Requirements for Release 1999".

[4]   Void.

[5]   3GPP TS 22.105: "Services & Service capabilities".

[6]   3GPP TS 24.008: "Mobile radio interface layer 3 specification; Core Network Protocols – Stage 3".

[7]   3GPP TS 23.207: "End-to-end QoS concept and architecture".

[8]   3GPP TS 23.008: "Organization of subscriber data".

[9]   3GPP TS 23.067: "enhanced Multi-Level Precedence and Pre-emption service (eMLPP) - Stage 2".

[10]   3GPP TS 03.60 (Release 1998): "Digital cellular telecommunications system (Phase 2+); General Packet Radio Service (GPRS); Service description; Stage 2 (Release 1998)".

[11]   3GPP TS 23.216: "Single Radio Voice Call Continuity (SRVCC); Stage 2".

3 Abbreviations

For the purpose of the present document, the following abbreviations apply:

3G	3rd Generation
AMR	Adaptive Multirate speech codec
ATM	Asynchronous Transfer Mode
BER	Bit Error Rate
BS	Bearer Service
CC	Call Control
CN	Core Network
CRC	Cyclic Redundancy Check
CS	Circuit Switched
DTX	Discontinuous Transmission

FDD	Frequency Division Duplex
FER	Frame Erasure Ratio
FTP	File Transfer Protocol
GERAN	GSM/EDGE Radio Access Network
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communication
IETF	Internet Engineering Task Force
IP	Internet Protocol
ISDN	Integrated Services Digital Network
MO	Mobile Originating Call
MPEG	Moving Pictures Expert Group
MT	Mobile Terminal
MTC	Mobile Terminated Call
NS	Network Service
PDP	Packet Data Protocol
PDU	Protocol Data Unit
PS	Packet Switched
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RA	Routing Area
RAB	Radio Access Bearer
RAN	Radio Access Network
RLC	Radio Link Control
RSVP	Resource Reservation Protocol
RT	Real Time
RTP	Real Time Transport Protocol
SAP	Service Access Point
SDU	Service Data Unit
SGSN	Serving GPRS Support Node
SLA	Service Level Agreement
SMS	Short Message Service
SVC	Switched Virtual Circuit
UDP	User Datagram Protocol
TBC	Token Bucket Counter
TDD	Time Division Duplex
TE	Terminal Equipment
TSPEC	Traffic Specification
UE	User Equipment
UMTS	Universal Mobile Telecommunication System
UTRA	UMTS Terrestrial Radio Access
UTRAN	UMTS Terrestrial Radio Access Network

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 attributes has to be as small as possible;

- derivation/definition of QoS attributes from the application requirements has to be simple;

- QoS attributes shall 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 attributes (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 attributes 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 attribute combinations have to have unambiguous meaning.

4.3   Technical Requirements for QoS

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

- UMTS QoS control mechanisms shall provide QoS attribute control on a peer to peer basis between UE 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 shall 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 attributes are needed to support asymmetric bearers;

- applications (or special software in UE 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 attributes during an active session;

- number of attributes should be kept reasonably low (increasing number of attributes, increase system complexity);

- user QoS requirements shall be satisfied by the system, including when change of SGSN within the Core Network occurs.

5 CS QoS in release 1999

For UMTS release '99 CS-CC, the QoS related bearer definitions of GSM (as defined in bearer capability information element, octet 6 and its extensions) are sufficient.

Based on the Bearer Capability information element the following services can be identified:

   a) **speech:** from the Information Transfer Capability (ITC) parameter;

   b) **data, non-transparent:** from the ITC and Connection element (CE) parameters;

   c) **data, transparent:** from the ITC and CE parameters.

For each of the above services, associated call control parameters, including the Bearer Capability information element, can be considered to define the UMTS bearer service.

The further mapping to Radio Access Bearer attributes is done according to the principles described in clause 8.

   NOTE: The mapping from GSM CC to UMTS RAB attributes is in the responsibility of CN WG1 and CN WG3.

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. An 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.
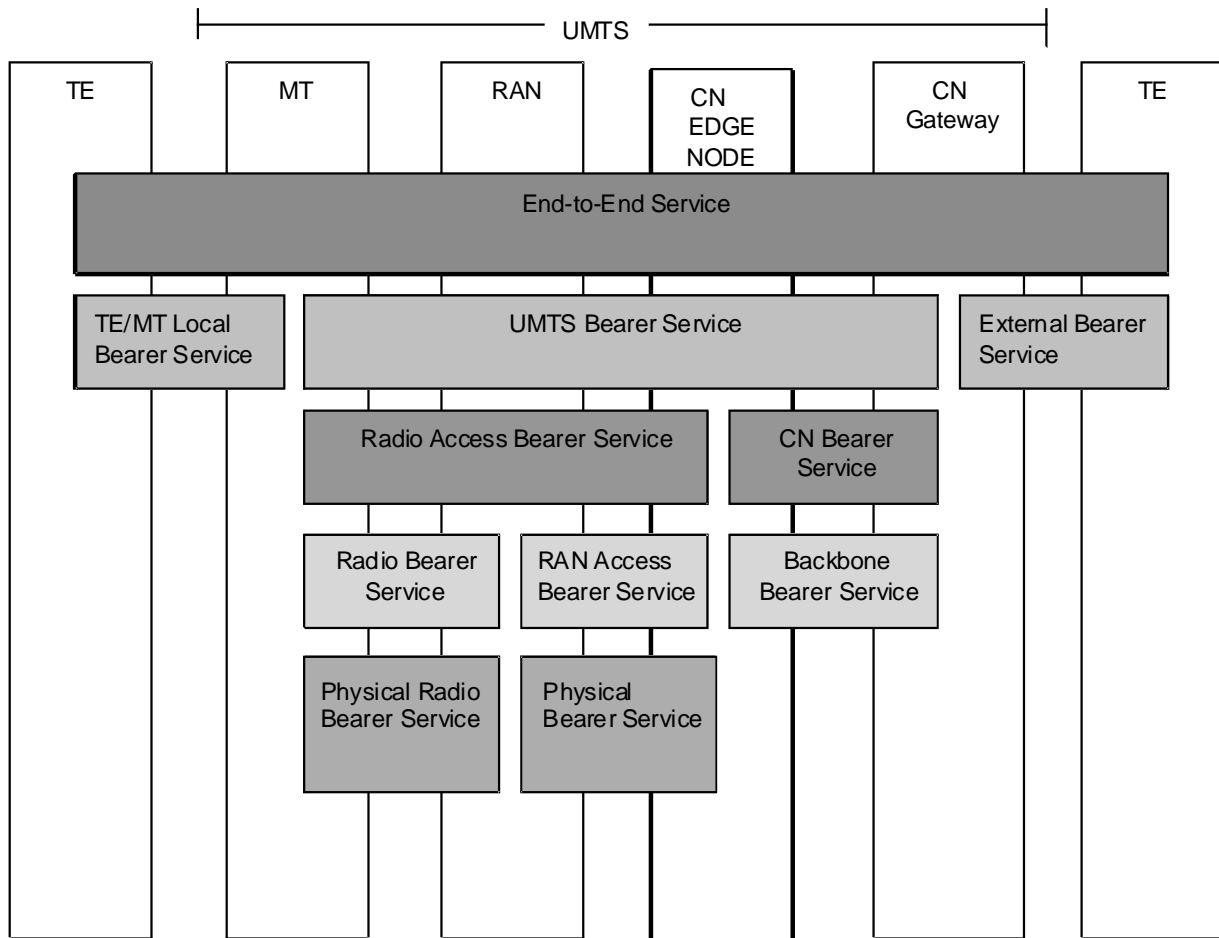
Figure 1: UMTS QoS Architecture

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 the present 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 the present 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 clause 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 reflects 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 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.

If unequal error protection shall be supported, it is provided by underlying Radio Bearer Services. In this case the payload of the user data SDU, transported by the Radio Access Bearer Service, shall conform to a SDU format defined with possible exact sizes and the payload bits statically structured per size. Each bit of the SDU payload belongs to a defined subflow. At Radio Access Bearer Service establishment, the exact SDU payload format and required reliability per subflow is signalled to RAN using standardised attributes (see clause 6.4.3).

In release 1999, unequal error protection for a Radio Access Bearer is only applicable for services using a codec integrated in the core network. This implies that UMTS Bearer service can not use the attribute SDU format information to define subflows and the payload bits of the SDUs will therefore be equally protected.

The Core Network Bearer Service of the UMTS core network connects the UMTS CN 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. The UMTS packet core network shall support different backbone bearer services for variety of QoS.

6.1.3   The Radio Bearer Service and the RAN Access Bearer Service

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

The Radio Bearer Service covers all the aspects of the radio interface transport. This bearer service is provided by the UTRAN FDD/TDD or the GERAN, which are not elaborated further in the present document.

To support unequal error protection, RAN and MT shall have the ability to segment/reassemble the user flows into the different subflows requested by the Radio Access Bearer Service. The segmentation/ reassemble is given by the SDU payload format signalled at Radio Access Bearer establishment. The Radio Bearer service handles the part of the user flow belonging to one subflow, according to the reliability requirements for that subflow.

The RAN Access Bearer Service together with the Physical Bearer Service provides the transport between RAN and CN. RAN Access bearer services for packet traffic shall provide different bearer services for variety of QoS. The RAN Access Bearer Service is provided by the Iu or the Gb Bearer Service.

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 clause is to give a comprehensive overview of functionality needed to establish, modify and maintain a UMTS Bearer Service with a specific QoS. The relations between the functions internal to the nodes are implementation specific. The allocation of these functions 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 attributes 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. The reference algorithm for traffic conditioning is described in Annex B. This reference algorithm should not be interpreted as a required implementation algorithm.

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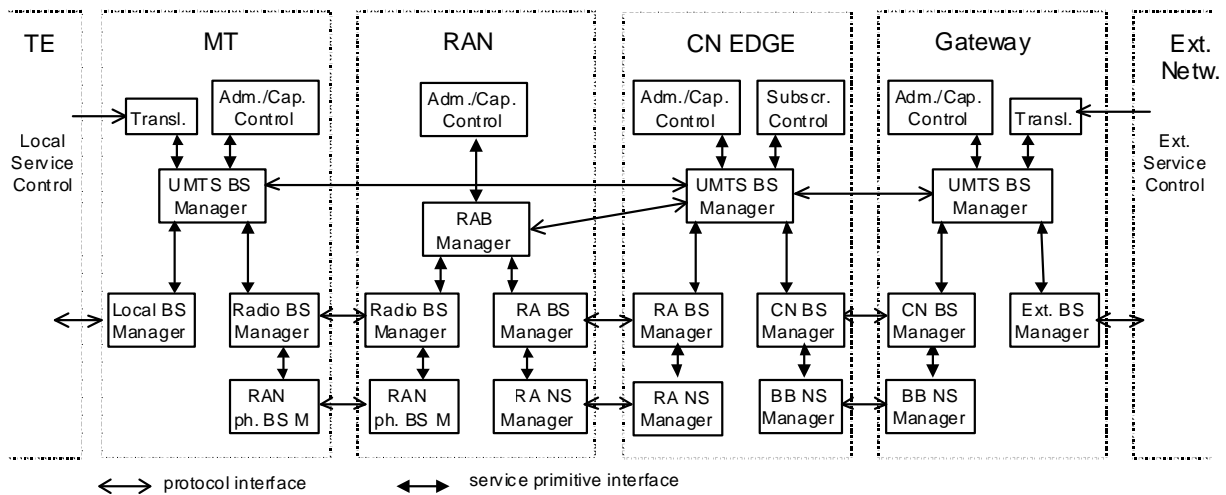
Figure 2: QoS management functions for UMTS bearer service in the control plane

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 RAN Access 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 RAN Access BS manager, its CN BS manager and the RAB manager in the RAN to provide the required services.

The RAB manager verifies with its admission/capability control whether the RAN supports the specific requested service and whether the required resources are available. It translates the RAB service attributes into radio bearer service and RAN Access bearer service attributes and requests the radio BS manager and the RAN Access 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, RAN Access 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.
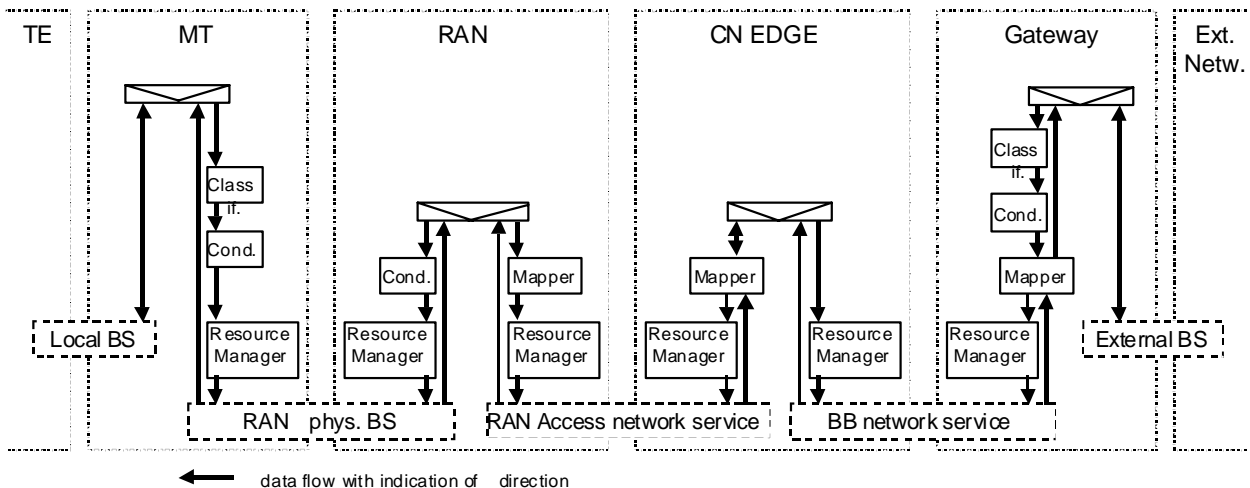
Figure 3: QoS management functions for the UMTS bearer service in the user plane

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; i.e. on a per PDP context basis. In addition, the traffic conditioner in the Gateway may provide conformance of the uplink and downlink user data traffic with QoS attributes related to the whole non-guaranteed bit-rate traffic of a UE for an APN (i.e. APN-AMBR). The packet oriented transport of the downlink data units from the external bearer service to the RAN and the buffering in the RAN may result in bursts of downlink data units not conformant with the UMTS BS QoS attributes. A traffic conditioner in the RAN forms this downlink data unit traffic according to the relevant QoS attributes. In addition, the traffic conditioner in the RAN may provide conformance of the uplink and downlink user data traffic (traffic) with QoS attributes related to the whole non-guaranteed bit-rate traffic of a UE (i.e. UE-AMBR).

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, also referred to as traffic 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 the QoS classes for UMTS.

There are four different QoS classes:

-      conversational class;

-      streaming class;

-      interactive class; and

-      background class.

The main distinguishing factor between these QoS 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.

However, these are only typical examples of usage of the traffic classes. There is in particular no strict one-to-one mapping between classes of service (as defined in TS 22.105 [5]) and the traffic classes defined in this TS. For instance, a service interactive by nature can very well use the Conversational traffic class if the application or the user has tight requirements on delay.

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 shall 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 shall 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 shall be preserved, although it does not have any requirements on low transfer delay.

The delay variation of the end-to-end flow shall 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 shall 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 shall 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	- Preserve time relation (variation) between information entities of the stream Conversational pattern (stringent and low delay)	- Preserve time relation (variation) between information entities of the stream	- Request response pattern - Preserve payload content	- 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 Attributes

NOTE: The discussion of UMTS bearer service attributes as well as radio access bearer attributes is still going on. Especially the bitrate attributes are under discussion and few comments have also been given to reliability attribute.

6.4.1 Asymmetric Bearers

Uni-directional and bi-directional bearer services shall be supported. For bi-directional bearer services, the attributes Maximum bitrate, and Guaranteed bitrate should be possible to set separately for uplink/downlink in order to support asymmetric bearers.

6.4.2 Sources of UMTS Bearer Service Attributes

UMTS bearer service attributes describe the service provided by the UMTS network to the user of the UMTS bearer service. A set of QoS attributes (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.

- If the UE requests or modifies a UMTS bearer and one or more of the QoS attributes are not specified by the UE by setting the attributes to 'subscribed', the SGSN shall assume a request of values as specified in the QoS profile in the UMTS subscription. If the UE sets the traffic class to 'subscribed', the SGSN shall assume a request for Interactive class. When the application in the UE requires streaming or conversational QoS, then the UE shall at least explicitly request the traffic class and should explicitly request the guaranteed bit rate and the maximum bit rate. For the rest of the QoS attributes, the network shall ensure that the negotiated QoS contains only values explicitly defined for the traffic class.

- 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.

6.4.3 UMTS Bearer Service Attributes

6.4.3.1 List of attributes

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 and to UMTS at a SAP within a period of time, divided by the duration of the period. The traffic is conformant with Maximum bitrate as long as it follows a token bucket algorithm where token rate equals Maximum bitrate and bucket size equals Maximum SDU size.

The conformance definition should not be interpreted as a required implementation algorithm. The token bucket algorithm is described in annex B.

The Maximum bitrate is the upper limit a user or application can accept or provide. All UMTS bearer service attributes may be fulfilled for traffic up to the Maximum bitrate depending on the network conditions.

[Purpose: Maximum bitrate can be used to make code reservations in the downlink of the radio interface. Its purpose is 1) to limit the delivered bitrate to applications or external networks with such limitations 2) to allow maximum wanted user bitrate to be defined for applications able to operate with different rates (e.g. applications with adapting codecs).]

Guaranteed bitrate (kbps)

Definition: guaranteed number of bits delivered by UMTS at a SAP within a period of time (provided that there is data to deliver), divided by the duration of the period. The traffic is conformant with the guaranteed bitrate as long as it follows a token bucket algorithm where token rate equals Guaranteed bitrate and bucket size equals Maximum SDU size.

The conformance definition should not be interpreted as a required implementation algorithm. The token bucket algorithm is described in annex B.

UMTS bearer service attributes, e.g. delay and reliability attributes, are guaranteed for traffic up to the Guaranteed bitrate. For the traffic exceeding the Guaranteed bitrate the UMTS bearer service attributes are not guaranteed.

[Purpose: Describes the bitrate the UMTS bearer service shall guarantee to the user or application. Guaranteed bitrate may be used to facilitate admission control based on available resources, and for resource allocation within UMTS.]

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]

Delivery order should be set to 'no' for PDP Type = 'IPv4' or 'IPv6'. The SGSN shall ensure that the appropriate value is set.

Maximum SDU size (octets)

Definition: the maximum SDU size for which the network shall satisfy the negotiated QoS.

[Purpose: The maximum SDU size is used for admission control and policing and/or optimising transport (optimized transport in for example the RAN may be dependent on the size of the packets). Handling by the network of packets larger than Maximum SDU size is implementation specific (e.g. they may be dropped or forwarded with decreased QoS).]

NOTE: The Maximum Transfer Unit (MTU) of the IP layer and the Maximum SDU Size have no relationship; in particular the GGSN should not perform IP fragmentation based on the Maximum SDU Size.

SDU format information (bits)

Definition: list of possible exact sizes of SDUs

[Purpose: RAN 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.]

SDU error ratio

Definition: Indicates the fraction of SDUs lost or detected as erroneous. SDU error ratio is defined only for conforming traffic.

NOTE 1: By reserving resources, SDU error ratio performance is independent of the loading conditions, whereas without reserved resources, such as in Interactive and Background classes, SDU error ratio is used as target value.

[Purpose: Used to configure the protocols, algorithms and error detection schemes, primarily within RAN.]

Residual bit error ratio

Definition: Indicates the undetected bit error ratio in the delivered SDUs. If no error detection is requested, Residual bit error ratio indicates the bit error ratio in the delivered SDUs.

[Purpose: Used to configure radio interface protocols, algorithms and error detection coding.]

Delivery of erroneous SDUs (y/n/-)

Definition: Indicates whether SDUs detected as erroneous shall be delivered or discarded.

NOTE 2: 'yes' implies that error detection is employed and that erroneous SDUs are delivered together with an error indication, 'no' implies that error detection is employed and that erroneous SDUs are discarded, and '-' implies that SDUs are delivered without considering error detection.

[Purpose: Used to decide whether error detection is needed and whether frames with detected errors shall be forwarded or not.]

Transfer delay (ms)

Definition: Indicates maximum delay for 95th percentile of the distribution of delay for all delivered SDUs during the lifetime of a bearer service, where delay for an SDU is defined as the time from a request to transfer an SDU at one SAP to its delivery at the other SAP.

[Purpose: relates to the delay tolerated by the application. In conjunction with the SDU error ratio attribute, care needs to be taken in deriving the value for the 95th percentile when an application desires, for example, that 99.9% of all transmitted packets are delivered within a certain time. This attribute allows RAN to set transport formats and ARQ parameters.]

NOTE 3: 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. The Allocation/Retention Priority attribute is a subscription attribute which is not negotiated from the mobile terminal, but the value might be changed either by the SGSN or the GGSN network element.

NOTE 4:   The addition of a user-controlled Allocation/Retention Priority attribute is for further study in future releases.

[Purpose: Priority is used for differentiating between bearers when performing allocation and retention of a bearer. In situations where resources are scarce, the relevant network elements can use the Allocation/Retention Priority to prioritize bearers with a high Allocation/Retention Priority over bearers with a low Allocation/Retention Priority when performing admission control.]

Source statistics descriptor ('speech'/'unknown')

Definition: specifies characteristics of the source of submitted SDUs.

Editor's note: The number of different source statistics descriptors that should be allowed is FFS.

[Purpose: Conversational speech has a well-known statistical behaviour (or the discontinuous transmission (DTX) factor). By being informed that the SDUs for a UMTS bearer are generated by a speech source, RAN, the SGSN and the GGSN and also the UE may, based on experience, calculate a statistical multiplex gain for use in admission control on the relevant interfaces.]

Signalling Indication (Yes/No)

Definition: Indicates the signalling nature of the submitted SDUs. This attribute is additional to the other QoS attributes and does not over-ride them. This attribute is only defined for the interactive traffic class. If signalling indication is set to 'Yes', the UE should set the traffic handling priority to '1'.

[Purpose: Signalling traffic can have different characteristics to other interactive traffic, e.g. higher priority, lower delay and increased peakiness. This attribute permits enhancing the RAN operation accordingly. An example use of the Signalling Indication is for IMS signalling traffic.]

NOTE: This indication is sent by the UE in the QoS IE.

Evolved Allocation/Retention Priority

Definition: Enhances the Allocation/Retention Priority attribute with an increased value range of the priority level and additional information about the pre-emption capability and the pre-emption vulnerability of the bearer. The pre-emption capability information defines whether a bearer with a lower priority level should be dropped to free up the required resources. The pre-emption vulnerability information whether a bearer is applicable for such dropping by a pre-emption capable bearer with a higher priority value.

6.4.3.2   Attributes discussed per traffic class

Conversational class

If the UMTS bearer carries speech service, **Source statistics descriptor** can be set, which allows UMTS to calculate a statistical multiplexing gain in core network, RAN and UE and use that for admission control.

The support for SRVCC requires conversational class and Source statistics descriptor set to speech only be used for IMS speech sessions in accordance to TS 23.216 [11].

NOTE: Triggering SRVCC will cause service interruption and/or inconsistent service experience when using conversational class and Source statistics descriptor set to speech for non-IMS services.

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 RAN without RLC re-transmissions. Hence, RAN 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 format information**. The SDU periodicity knowledge needed to operate in RLC transparent mode is obtained through dividing the largest defined SDU format by Maximum bitrate. This shall be considered when setting the attribute values in a service request.

The **Maximum SDU size** is only applicable if **SDU format information** is not specified and is used for admission control and policing and/or optimising transport. If **Maximum SDU size** is specified the SDU size is variable. If **SDU format information** is specified, with one or several possible sizes, each SDU shall exactly conform to one of the specified sizes. By using the **SDU error ratio**, **Residual bit error ratio** and **Delivery of erroneous SDUs** attribute, 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

If the UMTS bearer carries streaming speech service, **Source statistics descriptor** can be set, which allows UMTS to calculate a statistical multiplexing gain in core network, RAN and UE and use that for admission control.

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.

It shall be possible for Streaming bearers to be realised in RAN without RLC re-transmissions. Hence, RAN 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 format information**. The SDU periodicity knowledge needed to operate in RLC transparent mode is obtained through dividing the largest defined SDU format by Maximum bitrate. This shall be considered when setting the attribute values in a service request.

The **Maximum SDU size** is only applicable if **SDU format information** is not specified and is used for admission control and policing and/or optimising transport. If **Maximum SDU size** is specified the SDU size is variable. If **SDU format information** is specified, with one or several possible sizes, each SDU shall exactly conform to one of the specified sizes.

By using the **SDU error ratio**, **Residual bit error ratio** and **Delivery of erroneous SDUs** attribute, 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 RAN/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 given by **SDU error ratio**, **Residual bit error ratio** and **Delivery of erroneous SDUs**. Because there are no reserved resources

for interactive class, SDU error ratio should be used as a target value. SDU error ratio cannot be guaranteed under abnormal load conditions.

If the **Signalling Indication** is set, a statistical multiplexing gain and/or improvements in signalling speed may be obtained within the UTRAN.

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, given by **SDU error ratio**, **Residual bit error ratio** and **Delivery of erroneous SDUs** , is needed. Because there are no reserved resources for background class, SDU error ratio should be used as a target value. SDU error ratio cannot be guaranteed under abnormal load conditions.

6.4.3.3   UMTS bearer attributes: summary

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

Table 2: UMTS bearer attributes defined for each bearer traffic class

Traffic class	Conversational class	Streaming class	Interactive class	Background class
Maximum bitrate	X	X	X	X
Delivery order	X	X	X	X
Maximum SDU size	X	X	X	X
SDU format information	X	X		
SDU error ratio	X	X	X	X
Residual bit error ratio	X	X	X	X
Delivery of erroneous SDUs	X	X	X	X
Transfer delay	X	X		
Guaranteed bit rate	X	X		
Traffic handling priority			X	
Allocation/Retention priority	X	X	X	X
Source statistics descriptor	X	X		
Signalling indication			X	
Evolved Allocation/Retention priority	X	X	X	X

6.4.4   Radio Access Bearer Service Attributes

Radio Access Bearer Service Attributes shall be applied to both CS and PS domains.

6.4.4.1   List of attributes

Traffic class ('conversational', 'streaming', 'interactive', 'background')

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

[Purpose: By including the traffic class itself as an attribute, RAN can make assumptions about the traffic source and optimise 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 RAN and to RAN at a SAP within a period of time, divided by the duration of the period. The traffic is conformant with the Maximum bitrate as long as it follows a token bucket algorithm where token rate equals Maximum bitrate and bucket size equals Maximum SDU size.

The conformance definition should not be interpreted as a required implementation algorithm. The token bucket algorithm is described in annex B.

The Maximum bitrate is the upper limit a user or application can accept or provide. All RAB attributes may be fulfilled for traffic up to the Maximum bitrate depending on the network conditions.

> [Purpose: 1)   to limit the delivered bitrate to applications or external networks with such limitations, 2) to allow maximum wanted RAB bitrate to be defined for applications able to operate with different rates (e.g. applications with adapting codecs.)]

Guaranteed bitrate (kbps)

Definition: guaranteed number of bits delivered at a SAP within a period of time (provided that there is data to deliver), divided by the duration of the period. The traffic is conformant with the Guaranteed bitrate as long as it follows a token bucket algorithm where token rate equals Guaranteed bitrate and bucket size equals Maximum SDU size.

The conformance definition should not be interpreted as a required implementation algorithm. The token bucket algorithm is described in annex B.

RAB attributes, e.g. delay and reliability attributes, are guaranteed for traffic up to the Guaranteed bitrate. For the traffic exceeding the Guaranteed bitrate the RAB attributes are not guaranteed.

> [Purpose: Describes the bitrate the RAB shall guarantee to the user or application. Guaranteed bitrate may be used to facilitate admission control based on available resources, and for resource allocation within RAN.. 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.]

> NOTE: Delivery order should be set to 'no' for PDP Type = 'IPv4' or 'IPv6'.

Maximum SDU size (octets)

Definition: the maximum SDU size for which the network shall satisfy the negotiated QoS.

> [Purpose: The maximum SDU size is used for admission control and policing and/or optimising transport (optimized transport in for example the RAN may be dependent on the size of the packets). Handling by the network of packets larger than Maximum SDU size is implementation specific (e.g. they may be dropped or forwarded with decreased QoS).]

SDU format information (bits)

Definition: list of possible exact sizes of SDUs. If unequal error protection shall be used by a Radio Access Bearer service, SDU format information defines the exact subflow format of the SDU payload. SDU format information also supports definition of allowed subflow bitrates.

> NOTE 1: SDU format information is used by RAN to define which bits of the payload that belongs to each subflow. Exact syntax of SDU format information attribute is the task of RAN WG3.

> [Purpose: RAN needs SDU format 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. Moreover, in case of unequal error protection, RAN needs to know the exact format of SDU payload to be able to demultiplex the SDU onto different radio bearer services. When rate control is applied to services having a constant SDU size, e.g. CS data, the subflow bitrate is used to calculate the allowed inter PDU transmission interval (IPTI).]

SDU error ratio

Definition: Indicates the fraction of SDUs lost or detected as erroneous. SDU error ratio is defined only for conforming traffic. In case of unequal error protection., SDU error ratio is set per subflow and represents the error ratio in each subflow. SDU error ratio is only set for subflows for which error detection is requested.

>   NOTE 2:   By reserving resources, SDU error ratio performance is independent of the loading conditions, whereas without reserved resources, such as in Interactive and Background classes, SDU error ratio is used as target value.

>   [Purpose: Used to configure protocols, algorithms and error detection schemes, primarily within RAN.]

Residual bit error ratio

Definition: Indicates the undetected bit error ratio for each subflow in the delivered SDUs. For equal error protection, only one value is needed. If no error detection is requested for a subflow, Residual bit error ratio indicates the bit error ratio in that subflow of the delivered SDUs.

>   [Purpose: Used to configure radio interface protocols, algorithms and error detection coding. For services requiring unequal error protection, residual bit error ratio is given for each subflow.]

Delivery of erroneous SDUs (y/n/-)

Definition: Indicates whether SDUs with detected errors shall be delivered or not. In case of unequal error protection, the attribute is set per subflow.

>   NOTE 3:   'yes' implies that error detection is employed and that erroneous SDUs are delivered together with an error indication, 'no' implies that error detection is employed and that erroneous SDUs are discarded, and '-' implies that SDUs are delivered without considering error detection.

In case of unequal protection, different subflows may have different settings. Whenever there is a detected error in a subflow with 'no', the SDU is discarded, irrespective of settings in other subflows. For an SDU with multiple subflows with a 'yes' setting, there may be one error indication per subflow, or, if there is only one error indication per SDU, it indicates that an error was detected in at least one of these subflows. Exact definitions are the task of RAN3.

>   [Purpose: Used to decide whether error detection is needed and whether frames with detected errors shall be forwarded or discarded.]

Transfer delay (ms)

Definition: Indicates maximum delay for 95th percentile of the distribution of delay for all delivered SDUs during the lifetime of a bearer service, where delay for an SDU is defined as the time from a request to transfer an SDU at one SAP to its delivery at the other SAP.

>   [Purpose: permits the derivation of the RAN part of the total transfer delay for the UMTS bearer. In conjunction with the SDU error ratio attribute, care needs to be taken in deriving the value for the 95th percentile when an application desires, for example, that 99.9% of all transmitted packets are delivered within a certain time. This attribute allows RAN 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 RAN 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. For PS services the Allocation/Retention Priority attribute of the Radio access bearer is derived from the UMTS bearer service attributes Allocation/Retention Priority. Other attributes may be used in addition. For CS services the Allocation/Retention Priority attribute of the Radio access bearer is derived from the eMLPP priority level

attribute (TS 23.067 [9]) and/or the CS Allocation/Retention Priority attribute (TS 23.008 [8]) and/or the Mobile Station Category attribute (TS 23.008 [8]) (which is only available for subscribers in their home PLMN). Other attributes may be used in addition.

NOTE 4: The parameter is not negotiated from the mobile terminal. The addition of a user-controlled Allocation/Retention Priority attribute is for further study in future releases.

[Purpose: Priority is used for differentiating between bearers when performing allocation and retention of a bearer. In situations where resources are scarce, the relevant network elements can use the Allocation/Retention Priority to prioritize bearers with a high Allocation/Retention Priority over bearers with a low Allocation/Retention Priority when performing admission control.]

Source statistics descriptor ('speech'/'unknown')

Definition: specifies characteristics of the source of submitted SDUs.

[Purpose: Conversational speech has a well-known statistical behaviour (or the discontinuous transmission (DTX) factor). By being informed that the SDUs for a RAB are generated by a speech source, RAN may, based on experience, calculate a statistical multiplex gain for use in admission control on the radio and RAN Access interfaces.]

Signalling Indication (Yes/No)

Definition: Indicates the signalling nature of the submitted SDUs. This attribute is additional to the other QoS attributes and does not over-ride them.

[Purpose: Signalling traffic can have different characteristics to other interactive traffic, e.g. higher priority, lower delay and increased peakiness. This attribute permits enhancing the RAN operation accordingly. An example use of the Signalling Indication is for IMS signalling traffic.]

6.4.4.2   Attributes discussed per traffic class

Conversational class

If the RAB carries a speech service, **Source statistics descriptor** can be set, which allows RAN to calculate a statistical multiplexing gain on radio and RAN Access interfaces and use that for admission control.

The support for SRVCC requires conversational class and Source statistics descriptor set to speech only be used for IMS speech sessions in accordance to TS 23.216 [11].

NOTE: Triggering SRVCC will cause service interruption and/or inconsistent service experience when using conversational class and Source statistics descriptor set to speech for non-IMS services.

Unequal error protection can be supported in conversational class. In case unequal error protection is requested for a given RAB, the attributes Delivery of erroneous SDUs, Residual bit error ratio and SDU error ratio are specified per subflow. **Delivery of erroneous SDUs** determines whether error detection shall be used and, if so, whether SDUs with error in a certain subflow shall be delivered or not. **Residual bit error ratio** specifies the bit error ratio for undetected delivered bits. **SDU error ratio** specifies the fraction of SDUs with detected error in each subflow. It is only set for subflows for which error detection is requested.

In case of unequal error protection the payload of the user data SDU, transported by the Radio Access Bearer Service, shall conform to a SDU format defined with possible exact sizes. The payload bits are statically structured into subflows. The **SDU format information** attribute defines the exact subflow format of SDU payload.

RAN includes a rate control protocol, making it able of controlling the rate of sources requesting this, provided that they are periodic and that **SDU format information** is specified. RAN is allowed to control the rate between **Guaranteed bitrate** and **Maximum bitrate**. Each of these two rates shall correspond to an SDU format specified in **SDU format information**. For the case the SDU size is constant, as is the case for CS data, SDU format information may include a list of possible bitrates per subflow, to allow rate control of the subflows by change of inter PDU transmission interval (IPTI).

Streaming class

If the RAB carries streaming speech, **Source statistics descriptor** can be set, which allows RAN to calculate a statistical multiplexing gain on radio and RAN Access interfaces and use that for admission control.

Unequal error protection can be supported in streaming class. In case unequal error protection is requested for a given RAB, the attributes Delivery of erroneous SDUs, Residual bit error ratio and SDU error ratio are specified per subflow. **Delivery of erroneous SDUs** determines whether error detection shall be used and, if so, whether SDUs with error in a certain subflow shall be delivered or not. **Residual bit error ratio** specifies the bit error ratio for undetected delivered bits. **SDU error ratio** specifies the fraction of SDUs with detected error in each subflow. It is only set for subflows for which error detection is requested.

In case of unequal error protection the payload of the user data SDU, transported by the Radio Access Bearer Service, shall conform to a SDU format defined with possible exact sizes. The payload bits are statically structured into subflows. The **SDU format information** attribute defines the exact subflow format of SDU payload.

RAN includes a rate control protocol, making it able of controlling the rate of sources requesting this, provided that they are periodic and that **SDU format information** is specified. RAN is allowed to control the rate between **Guaranteed bitrate** and **Maximum bitrate**. Each of these two rates shall correspond to an SDU format specified in **SDU format information**. For the case the SDU size is constant, as is the case for CS data, SDU format information may include a list of possible bitrates per subflow, to allow rate control of the subflows by change of inter PDU transmission interval (IPTI).

Other classes

The RAB attribute sets and their use in, interactive and background classes are identical to those of UMTS bearer services (clause 6.4.2.2).

6.4.4.3   Radio Access Bearer attributes: summary

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

Table 3: Radio Access Bearer attributes defined for each bearer traffic class

Traffic class	Conversational class	Streaming class	Interactive class	Background class
Maximum bitrate	X	X	X	X
Delivery order	X	X	X	X
Maximum SDU size	X	X	X	X
SDU format information	X	X		
SDU error ratio	X	X	X	X
Residual bit error ratio	X	X	X	X
Delivery of erroneous SDUs	X	X	X	X
Transfer delay	X	X		
Guaranteed bit rate	X	X		
Traffic handling priority			X	
Allocation/ Retention priority	X	X	X	X
Source statistics descriptor	X	X		
Signalling Indication			X	

6.4.5   Radio Bearer Service Attributes

NOTE: Defining the radio bearer service attributes is a task for RAN WG2.

6.4.6   RAN Access Bearer Service Attributes

The RAN Access-Bearer Service together with the Physical Bearer Service provides the transport between RAN and CN. RAN Access bearer services for packet traffic shall provide different bearer services for variety of QoS. It is operators' option which of QoS capabilities in Frame Relay layer, in IP layer or in ATM layer is used. For IP based RAN Access bearer services, Differentiated Services defined by IETF shall be used. If operator choose ATM-SVC as an internal dedicated transport bearer, inter operation with IP based networks will be based on Differentiated Services. The mapping from UMTS QoS classes to Diffserv codepoints will be controlled by the operator. The mapping depends on bandwidth and provisioning of resources among the different Diffserv classes which the operators control to satisfy their cost and performance requirements. Interoperability between operators will be based on the use of service level agreements (SLAs) which are an integral part of the Diffserv Architecture.

6.4.7   Core Network Bearer Service Attributes

The UMTS packet core network shall support different backbone bearer services for variety of QoS. It is operators' option which of QoS capabilities in IP layer or QoS capabilities in ATM layer is used. For the IP based backbone, Differentiated Services defined by IETF shall be used. If operator choose ATM-SVC as an internal dedicated transport bearer, interoperation with IP based backbone networks will be based on Differentiated Services. The mapping from UMTS QoS classes to Diffserv codepoints will be controlled by the operator. The mapping depends on bandwidth and provisioning of resources among the different Diffserv classes which the operators control to satisfy their cost and performance requirements. Interoperability between operators will be based on the use of service level agreements (SLAs) which are an integral part of the Diffserv Architecture.

6.5   Attribute Value Ranges

For UMTS Bearer service and Radio Access Bearer services a list of finite attribute values or the allowed value range is defined for each attribute. The value list/value range defines the values that are possible to be used for an attribute considering every possible service condition. When a service is defined as a combination of attributes, further limitations may apply; for example the shortest possible delay may not be possible to use together with the lowest possible SDU error ratio. Service requirements, i.e. required QoS and performance for a given UMTS service is defined in the service requirement specifications TS 22.105 [5]. The aspect of future proof coding of attributes in protocol specifications is not considered in the defined value list/value range tables.

6.5.1   Ranges of UMTS Bearer Service Attributes

The following table lists the value ranges of the UMTS bearer service attributes. The value ranges reflect the capability of UMTS network.

Table 4: Value ranges for UMTS Bearer Service Attributes

Traffic class	Conversational class	Streaming class	Interactive class	Background class
Maximum bitrate (kbps)	<= 256 000 (2)	<= 256 000 (2)	<= 256 000 (2)	<= 256 000 (2)
Delivery order	Yes/No	Yes/No	Yes/No	Yes/No
Maximum SDU size (octets)	<=1 500 or 1 502 (4)	<=1 500 or 1 502 (4)	<=1 500 or 1 502 (4)	<=1 500 or 1 502 (4)
SDU format information	(5)	(5)		
Delivery of erroneous SDUs	Yes/No/- (6)	Yes/No/- (6)	Yes/No/- (6)	Yes/No/- (6)
Residual BER	$5*10^{-2}$, 10^{-2}, $5*10^{-3}$, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}	$5*10^{-2}$, 10^{-2}, $5*10^{-3}$, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}	$4*10^{-3}$, 10^{-5}, $6*10^{-8}$ (7)	$4*10^{-3}$, 10^{-5}, $6*10^{-8}$ (7)
SDU error ratio	10^{-2}, $7*10^{-3}$, 10^{-3}, 10^{-4}, 10^{-5}	10^{-1}, 10^{-2}, $7*10^{-3}$, 10^{-3}, 10^{-4}, 10^{-5}	10^{-3}, 10^{-4}, 10^{-6}	10^{-3}, 10^{-4}, 10^{-6}
Transfer delay (ms)	100 – maximum value	300 (8) – maximum value		
Guaranteed bit rate (kbps)	<= 256 000 (2)	<= 256 000 (2)		
Traffic handling priority			1,2,3 (9)	
Allocation/Retention priority	1,2,3	1,2,3	1,2,3	1,2,3
Source statistic descriptor	Speech/unknown	Speech/unknown		
Signalling Indication			Yes/No (9)	
Evolved Allocation/Retention priority - Priority Level - Pre-emption Capability - Pre-emption Vulnerability	1-15 Yes/No Yes/No	1-15 Yes/No Yes/No	1-15 Yes/No Yes/No	1-15 Yes/No Yes/No

1) Void.

2) The granularity of the bitrate varies across the whole range from 1kbps steps up to 2 Mbps steps.

3) Void.

4) In case of PDP type = PPP, maximum SDU size is 1502 octets. In other cases, maximum SDU size is 1 500 octets.

5) Definition of possible values of exact SDU sizes for which RAN can support transparent RLC protocol mode, is the task of RAN WG3.

6) If *Delivery of erroneous SDUs* is set to 'Yes' error indications can only be provided on the MT/TE side of the UMTS bearer. On the CN Gateway side error indications can not be signalled outside of UMTS network in release 1999.

7) Values are derived from CRC lengths of 8, 16 and 24 bits on layer 1.

8) If the UE requests a transfer delay value lower than the minimum value, this shall not cause the network (SGSN and GGSN) to reject the request from the UE. The network may negotiate the value for the transfer delay.

9) If signalling indication is set to 'Yes', the UE should set the traffic handling priority to '1'.

6.5.2 Ranges of Radio Access Bearer Service Attributes for UTRAN and for GERAN

The following table lists the value ranges of the radio access bearer service attributes for UTRAN and for GERAN. The value ranges reflect the capability of both UTRAN and GERAN.

Table 5: Value ranges for Radio Access Bearer Service Attributes for UTRAN and for GERAN

Traffic class	Conversational class	Streaming class	Interactive class	Background class
Maximum bitrate (kbps)	<= 256 000 (2) (7)	<= 256 000 (2) (7)	<= 256 000 (2) (7)	<= 256 000 (2) (7)
Delivery order	Yes/No	Yes/No	Yes/No	Yes/No
Maximum SDU size (octets)	<=1 500 or 1 502 (4)	<=1 500 or 1 502 (4)	<=1 500 or 1 502 (4)	<=1 500 or 1 502 (4)
SDU format information (1)	(5)	(5)		
Delivery of erroneous SDUs	Yes/No/-	Yes/No/-	Yes/No/-	Yes/No/-
Residual BER	$5*10^{-2}$, 10^{-2}, $5*10^{-3}$, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}	$5*10^{-2}$, 10^{-2}, $5*10^{-3}$, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}	$4*10^{-3}$, 10^{-5}, $6*10^{-8}$ (6)	$4*10^{-3}$, 10^{-5}, $6*10^{-8}$ (6)
SDU error ratio	10^{-2}, $7*10^{-3}$, 10^{-3}, 10^{-4}, 10^{-5}	10^{-1}, 10^{-2}, $7*10^{-3}$, 10^{-3}, 10^{-4}, 10^{-5}	10^{-3}, 10^{-4}, 10^{-6}	10^{-3}, 10^{-4}, 10^{-6}
Transfer delay (ms)	80 – maximum value	250 – maximum value		
Guaranteed bit rate (kbps)	<= 256 000 (2) (7)	<= 256 000 (2) (7)		
Traffic handling priority			1,2,3	
Allocation/Retention priority (1) - Priority Level - Pre-emption Capability - Pre-emption Vulnerability	1,2, ..., 15 Yes/No Yes/No	1,2, ..., 15 Yes/No Yes/No	1,2, ..., 15 Yes/No Yes/No	1,2, ..., 15 Yes/No Yes/No
Source statistic descriptor	Speech/unknown	Speech/unknown		
Signalling Indication			Yes/No	

1) This parameter is limited to the values 1, 2 and 3 for GERAN when the Gb Bearer Service is used and the pre-emption capability and vulnerability information is not used.

2) The granularity of the bitrate varies across the whole range from 1kbps steps up to 2 Mbps steps.

3) Void.

4) In case of PDP type = PPP, maximum SDU size is 1502 octets. In other cases, maximum SDU size is 1 500 octets.

5) Definition of possible values of exact SDU sizes for which UTRAN can support transparent RLC protocol mode, is the task of RAN WG3.

6) Values are derived from CRC lengths of 8, 16 and 24 bits on layer 1.

7) In case of GERAN the highest bitrate value is 473.6 kbps.

6.6 Bearer unrelated QoS Attributes

APN-Aggregated Maximum Bit-rate (kbps)

Definition: The APN-AMBR limits the aggregate bit rate that can be expected to be provided across all Non GBR PDP contexts of the same APN (e.g. excess traffic may get discarded by a rate shaping function). Each of those Non GBR PDP contexts could potentially utilize the entire APN AMBR, e.g. when the other Non GBR PDP contexts do not carry any traffic. GBR PDP contexts are outside the scope of APN AMBR. The GGSN enforces the APN AMBR in downlink. Enforcement of APN AMBR in uplink may be done in the UE and additionally in the GGSN.

NOTE: All simultaneous active PDP contexts of a UE that are associated with the same APN need to be connected to the same GGSN.

[Purpose: The APN-AMBR can be used for a differentiation of users regarding the bit rate they can experience.]

UE Aggregated Maximum Bit-rate (kbps)

Definition: The UE AMBR limits the aggregate bit rate that can be expected to be provided across all Non GBR PDP contexts of a UE (e.g. excess traffic may get discarded by a rate shaping function). Each of those Non GBR PDP contexts could potentially utilize the entire UE AMBR, e.g. when the other Non GBR PDP contexts do not carry any

traffic. GBR (real-time) PDP contexts are outside the scope of UE AMBR. The RAN enforces the UE AMBR in uplink and downlink.

[Purpose: The UE AMBR can be used for limiting the overall bit rate a UE can experience.]

7 Void

8 QoS Attribute Mapping

NOTE: This clause shall contain information of attribute mapping i.e. how attribute in different levels of QoS (from external world and within UMTS network) shall be mapped. Current clause division is based on an assumption that levels of QoS presented in clause 6, but they are open to discussions.

8.1 From Application Attributes to UMTS Bearer Service Attributes

NOTE: This is an operator and/or implementation issue.

8.2 From UMTS Bearer Service Attributes to Radio Access Bearer Service Attributes

When establishing a UMTS bearer and the underlying Radio Access Bearer for support of a service request, some attribute on UMTS level does typically not have the same value as corresponding attribute on Radio Access Bearer level. For example requested transfer delay for the UMTS bearer shall typically be larger that the requested transfer delay for the Radio Access Bearer, as the transport through the core network will use a part of the acceptable delay.

For the following attributes/settings the attribute value for the UMTS bearer will normally be the same as the corresponding attribute value for the Radio Access Bearer:

- maximum bitrate;

- delivery order;

- delivery of erroneous SDUs;

NOTE 1: If *Delivery of erroneous SDUs* is set to 'Yes' the handling of error indications on UMTS Bearer level and Radio Access Bearer level differs. Error indications can only be provided on the MT/TE side of the UMTS bearer. On the CN Gateway side error indications can not be signalled outside of UMTS network in release 1999. Error indications can be provided on both end-points of the Radio Access Bearer.

- guaranteed bit rate;

- traffic handling priority;

- maximum SDU size;

- SDU format information.

NOTE 2: List of exact sizes of SDU's shall be the same, exact format of SDU payload does not exist on UMTS Bearer level.

For the following attributes the attribute value for the UMTS bearer will normally not be the same as the corresponding attribute value for the Radio Access Bearer. The relation between the attribute values for UMTS Bearer service and Radio Access Bearer service is implementation dependent and depends for example on network dimensioning.

- **Residual BER** for Radio Access Bearer service shall be reduced with the bit errors introduced in the core network, by Core Network Bearer service.

- **SDU error ratio** for Radio Access Bearer service shall be reduced with the errors introduced in the core network, by Core Network Bearer service.

- **Transfer delay** for Radio Access Bearer service shall be reduced with the delay introduced in the core network, e.g. on transmission links or in a codec resident in the Core Network.

For the following attribute the value range for the UMTS bearer is not the same as the corresponding attribute value ranges for the Radio Access Bearer. The value for the Radio Access Bearer service attribute shall be derived from the values of one or more UMTS bearer service attributes:

- For PS services the Allocation/Retention Priority shall be derived from the UMTS bearer service attributes Allocation/Retention Priority. Other attributes may be used in addition. For CS services the Allocation/Retention Priority shall be derived from the eMLPP priority level attribute and/or the CS Allocation/Retention Priority attribute and/or the Mobile Station Category attribute (TS 23.008 [8]) (which is only available for subscribers in their home PLMN). Other attributes may be used in addition.

NOTE 3: The usage of other attributes can ensure that e.g. emergency calls would receive an appropriate RAB Allocation/Retention Priority.

NOTE 4: To ensure backwards compatibility different values of an UMTS bearer service attribute can be mapped to the same value of the RAB Allocation/Retention Priority attribute if necessary.

NOTE 5: From Release 9 onwards, the UMTS bearer service attribute Evolved Allocation/Retention Priority can be mapped directly to the of the RAB Allocation/Retention Priority attribute.

The following attributes/settings only exist on the Radio Access Bearer level:

- **SDU format information** - exact format of SDU payload is retrieved from the codec integrated in the core network.

- **Source statistics descriptor** is set to speech if the Radio Access Bearer transports compressed speech generated by the codec integrated in the core network.

8.3 From UMTS Bearer Service Attributes to CN Bearer Service Attributes

NOTE: This is operator's choice.

9 Interworking

The model for the UMTS QoS classes and attributes 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 clause presents the most common technologies that UMTS shall be capable to interwork with.

9.1 UMTS-GSM CS/GPRS

9.1.1 UMTS-GSM CS

The mapping between UMTS-GSM CS is based on GSM CS mechanisms and CC parameters.

9.1.1.1   Handover from UMTS to GSM

In case a UMTS call is set up in the CN, the BC IEs are mapped into QoS RAB attributes at call setup.

If the CN has to perform a handover towards GSM, the non-anchor MSC needs to perform an assignment based on GSM specific traffic channel attributes.

As the BSSMAP protocol is used over the E-interface and as no appropriate procedure exists to map QoS attributes into BSSMAP parameters, the anchor MSC shall map BC IEs into GSM traffic channel parameters, according to existing GSM procedures for call setup.

This requires that the BC IE is coded according to GSM protocol requirements, i.e. all those parameters not applicable to UMTS should nevertheless be correctly specified by the UE in order to perform a handover to GSM according the above specified principles.

9.1.1.2   Handover from GSM to UMTS

In case a GSM call is set up in the CN, the BC IEs are mapped into channel type parameters at call setup.

If the CN has to perform a handover towards UMTS, the non-anchor MSC needs to perform an assignment based on UMTS specific radio access bearer attributes.

As the BSSMAP protocol is used over the E-interface, the non-anchor MSC shall use the received Channel Type parameter (e.g. 'speech or data indicator', the type of data service (transparent/non-transparent) and user rate) to derive the QoS RAB attributes.

9.1.2   UMTS-GPRS

This section covers primarily the mapping of QoS attributes that are necessary across standardised interfaces. In addition to these, there are cases when mapping of QoS attributes are needed internal to a node.

GPRS Release 99 (R99) QoS attributes shall be equivalent to the UMTS QoS Attributes. For interworking purposes between different releases, mapping rules between GPRS Release 97/98 (R97/98) and GPRS Release 99 (R99) as well as UMTS are defined. Mapping shall occur whenever the UE, the SGSN, the GGSN and the HLR nodes are of different releases R97/98 or R99. The mapping is required in PDP context activation and modification procedures and when a R99 HLR Insert Subscriber Data towards a R97/98 SGSN.

It is not within the scope of the present document to determine if any value combinations of attribute values can not be supported. This means that complete mapping rules are defined here, and if the user requests a QoS profile which the network is not able to support (e.g. a low delay and a high reliability), the decision if such a attribute combination can be supported is left to admission control functionality within the PDP context activation procedure, and the QoS for such a profile may be renegotiated by the network based on the available resources.

The overall principle for the mapping between two profiles is that the two profiles, applied in their respective network releases, give the same or at least similar QoS. The GPRS R97/98 equipment will not be able to provide realtime service corresponding to the R99 conversational and streaming traffic classes. Therefore, the mapping is always to the non-realtime interactive and background traffic classes.

The R97/R98 QoS attributes are described in TS 03.60 Release 1998 [10].

9.1.2.1   General rules

Air interface Session Management and GTP messages of R99 shall contain the R99 attributes as an extension of the R97/98 QoS Information Element thus unnecessary mapping can be avoided. When a R97/98 MS is visiting a GPRS R99 or UMTS SGSN and the GGSN is of R97/98 or R99, the visited SGSN shall not perform any mapping of QoS attributes. In case of GGSN R99, the GTP version 1 (R99) QoS profile only contains the R97/98 QoS attributes. It can be noted that for this PDP Context a Traffic Flow Template (TFT) can not be requested.

When a R99 UE is visiting a GPRS R99 or UMTS SGSN (or serving PLMN) and the GGSN (or home PLMN) is of R97/98, the visited SGSN (or visited PLMN) shall be capable of providing bearers having QoS support according to R99. When a PDP Context is activated (mobile or network initiated) mapping takes place in the serving SGSN.

For MS initiated PDP Context Activations as well as network initiated PDP Context Activations, the home R97/98 GGSN will respond to the activation request by returning a the QoS Negotiated Profile, which contain the accepted and changed R97/98 attributes. A mapping of the changed attributes into R99 attributes will be done in serving SGSN and signalled to the UE in the Activate PDP Context Accept message.

It is a general mapping rule that returned and unchanged attributes during negotiation procedures shall not be mapped a second time by serving SGSN, i.e. the unchanged R99 attributes received in the Create PDP Context Response message will be sent to UE in QoS Negotiated Profile of the Activate PDP Context Accept message.

MAP message of R99 shall also contain the R99 attributes as an extension of the R97/98 QoS Information Element when Insert Subscriber Data message is sent to a R99 SGSN. In the case when a R99 HLR send a Insert Subscriber Data message to a R97/98 SGSN, the message shall contain the R97/98 QoS attributes. A R99 SGSN shall use the R99 attributes of subscribed QoS profile when a R99 UE requests to use subscription data in the PDP Context Activation. The R99 SGSN shall use the R97/98 attributes of subscribed QoS profile when a R97/98 MS requests to use subscription data in the PDP Context Activation.

9.1.2.2   Determining R99 attributes from R97/98 attributes

This mapping is applicable in the following cases:

- hand over of PDP Context from GPRS R97/98 SGSN to GPRS R99 or UMTS SGSN;

- PDP Context Activation in a serving R99 SGSN with a R97/98 GGSN. When GGSN respond to the PDP Context Activation, mapping of the changed R97/98 QoS attributes received from the GGSN to R99 QoS attributes is performed in the serving SGSN.

This mapping is also applicable if a R99 UE allows an application to request a PDP Context Activation with R97/98 QoS attributes, e.g. via AT command.

Table 6: Rules for determining R99 attributes from R97/98 attributes

Resulting R99 Attribute		Derived from R97/98 Attribute	
Name	Value	Value	Name
Traffic class	Interactive	1, 2, 3	Delay class
	Background	4	
Traffic handling priority	1	1	Delay class
	2	2	
	3	3	
SDU error ratio	10^{-6}	2	Reliability class
	10^{-4}	3	
	10^{-3}	4, 5	
Residual bit error ratio	10^{-5}	2, 3, 4	Reliability class
	$4*10^{-3}$	5	
Delivery of erroneous SDUs	'no'	2, 3, 4	Reliability class
	'yes'	5	
Maximum bitrate [kbps]	8	1	Peak throughput class
	16	2	
	32	3	
	64	4	
	128	5	
	256	6	
	512	7	
	1024	8	
	2048	9	
Allocation/Retention priority	1	1	Precedence class
	2	2	
	3	3	
Delivery order	yes'	yes'	Reordering Required (Information in the SGSN and the GGSN PDP Contexts)
	'no'	'no'	
Maximum SDU size	1 500 octets	(Fixed value)	
Priority Level of Evolved Allocation/Retention priority	1 to H	1	Precedence class
	H+1 to M	2	
	M+1 to 15	3	

NOTE 1: As the allocation/retention priority attribute is not available in the UE (see 6.4.4.1) the mapping of the allocation/retention priority attribute is not relevant for the UE.

NOTE 2: The values of H (high priority) and M (medium priority) can be set according to operator requirements to ensure proper treatment of users with higher priority level information. The minimum value of H is 1. The minimum value of M is H+1.

As the reordering required attribute is not available in the MS the MS should set the R99 delivery order attribute to the value "no" for PDP Type = "IPv4" or "IPv6", otherwise the MS shall set the delivery order attribute to the value "subscribed" (see TS 24.008 [6]).

9.1.2.3   Determining R97/98 attributes from R99 attributes

This mapping is applicable in the following cases:

- PDP Context is handed over from GPRS R99 or UMTS to GPRS R97/98;

- when a R99 UE perform a PDP Context Activation in a serving R99 SGSN while the GGSN is of R97/98. In this case the SGSN shall perform mapping of the R99 QoS attributes to the R97/98 QoS attributes;

- a R99 HLR may need to map the stored subscribed QoS attributes in the HLR subscriber data to R97/98 QoS attributes that are going to be sent in the Insert Subscriber Data message from the R99 HLR to the R97/98 and R99 SGSN. It is an implementation issue if the R97/98 QoS attributes are stored in the HLR in addition to the R99 QoS attributes;

- a R99 UE (except UMTS only UE) receives a request for a PDP Context Activation with R99 QoS attributes, e.g. via AT command.

Table 7: Rules for determining R97/98 attributes from R99 attributes

Resulting R97/98 Attribute		Derived from R99 Attribute	
Name	Value	Value	Name
Delay class	1	conversational	Traffic class
	1	streaming	Traffic class
	1	Interactive	Traffic class
		1	Traffic handling priority
	2	Interactive	Traffic class
		2	Traffic handling priority
	3	Interactive	Traffic class
		3	Traffic handling priority
	4	Background	Traffic class
Reliability class	2	$<= 10^{-5}$	SDU error ratio
	3	$10^{-5} < x <= 5*10^{-4}$	SDU error ratio
	4	$> 5*10^{-4}$	SDU error ratio
		$<= 2*10^{-4}$	Residual bit error ratio
	5	$> 5*10^{-4}$	SDU error ratio
		$> 2*10^{-4}$	Residual bit error ratio
Peak throughput class	1	< 16	Maximum bitrate [kbps]
	2	16 <= x < 32	
	3	32 <= x < 64	
	4	64 <= x < 128	
	5	128 <= x < 256	
	6	256 <= x < 512	
	7	512 <= x < 1024	
	8	1024 <= x < 2048	
	9	>= 2048	
Precedence class	1	1	Allocation/retention priority
	2	2	
	3	3	
Mean throughput class	Always set to 31	-	
Reordering Required (Information in the SGSN and the GGSN PDP Contexts)	yes'	yes'	Delivery order
	'no'	'no'	
Precedence class	1	1	Priority Level of the Evolved Allocation/retention priority
	2	H+1	
	3	M+1	

As the allocation/retention priority attribute is not available in the UE (see 6.4.4.1) the UE shall set the R97/98 precedence class attribute to the value "subscribed" (see TS 24.008 [6]).

NOTE 1: In the case of asymmetric bearers, the higher value of the maximum bitrate attributes for downlink and uplink is selected and used for the maximum bitrate value.

NOTE 2: The values of H (high priority) and M (medium priority) can be set according to operator requirements to ensure proper treatment of users with higher priority level information. The minimum value of H is 1. The minimum value of M is H+1.

9.2 UMTS-PSTN

PSTN does not have QoS mechanisms thus QoS attribute interworking/mapping is not needed.

NOTE: The details are to be solved by CN WG3.

9.3 UMTS-ISDN

ISDN does not have QoS mechanisms thus QoS attribute 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.

NOTE: The details are to be solved by CN WG3.

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 attributes. 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 clause.

IP based QoS models shall be supported for PDP contexts, meaning both Integrated Services (IntServ) signalled by RSVP [RFC 2205] and Differentiated Services (6-bit QoS attribute 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. The way application level QoS and UMTS level QoS interact is detailed in TS 23.207 [7].

Annex A (informative):
Error resilience in real-time packet multimedia payloads

A.1 Introduction

This annex provides some basic information with respect to the error resilience of different encoded media streams when considering the support of unequal error protection for real-time packet multimedia services. It provides some indicative figures for the residual bit error rates that could be tolerated by audio-visual H.323 payloads in a 3G environment.

H.323 employs the H.225.0 packetisation scheme, which in turn uses UDP/IP and RTP to transport each media stream. The structure of an H.323 packet is shown in figure A.1.
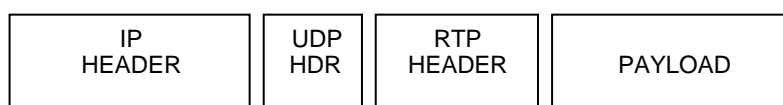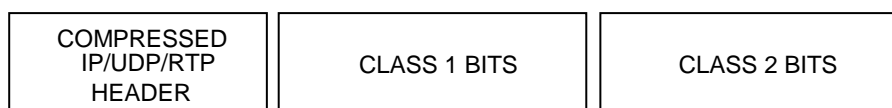
IP HEADER	UDP HDR	RTP HEADER	PAYLOAD

Figure A.1: Structure of H.323 packet

COMPRESSED IP/UDP/RTP HEADER	CLASS 1 BITS	CLASS 2 BITS

Figure A.2: Structure of compressed H.323 packet.
Class 1 bits can tolerate medium BER; Class 2 bits can tolerate high BER

It is assumed that some elements of the H.323 header information, which comprises the IP, UDP and RTP headers, can be compressed. It is also assumed that this information will require reliable transmission, such that any errors in the header will result in the loss of the complete H.323 packet. However, for real-time multimedia streams that cannot accommodate a large delay (and therefore packet retransmission), codecs can be used that are tolerant to residual bit errors.

This annex highlights the error resilience of audio and visual codecs, and provide some example tolerance figures for media streams of the type that are likely to comprise H.323 payloads.

A.1.1 Factors affecting error resilience

Specific error resilience figures will depend on a number of factors, including:

- the media type;

- the quality of service (QoS) required;

- the specific codec used.

Media streams may also be sub-divided into different classes on the basis of bit error sensitivity as shown in figure A.2. In some cases the most sensitive bits may be protected by in-band checksum information. It should also be noted that, in addition to the effect of residual bit errors in the media stream, the QoS will be further degraded by packet loss due to errors in the H.323 header.

A.2 Example figures

The following values are indicative of the QoS attributes required by audio and video media streams, including bit error rates (BER) and frame erasure rates (FER).

For the purposes of example, figures are provided for the AMR speech codec and the MPEG-4 video codec.

AMR speech codec payload

Bit rate: 4,75 - 12,2 kbit/s

Delay: end-to-end delay not to exceed 100ms (codec frame length is 20ms)

BER: 10^{-4} for Class 1 bits

10^{-3} for Class 2 bits

for some applications, a higher BER class ($\sim 10^{-2}$) might be feasible.

FER < 0,5% (with graceful degradation for higher erasure rates)

MPEG-4 video payload:

Bit rate: variable, average rate scalable from 24 to 128 kbit/s and higher

Delay: end-to-end delay between 150 and 400ms

video codec delay is typically less than 200 ms

BER: 10^{-6} - no visible degradation

10^{-5} - little visible degradation

10^{-4} - some visible artefacts

$> 10^{-3}$ - limited practical application

Packet loss rate FFS

Data and control:

Data (data refers to other types than audio and video e.g. file transfers, shared whiteboard) and control information shall be transmitted reliably (i.e. residual bit errors should result in a lost packet).

Annex B (normative):
Reference Algorithm for Conformance Definition of Bitrate

The annex shows a reference algorithm for the conformance definition of bitrate. This may be used for traffic contract between UMTS bearers and external network/user equipment. It should be noted that the reference algorithm will never imply a particular implementation for the traffic conditioner.

The algorithm is well known as "Token Bucket Algorithm" which has been described in IETF. Here, "tokens" represents the allowed data volume, for example in byte. "Tokens" are given at a constant "token rate" by a traffic contract, are stored temporarily in a "token bucket", and are consumed by accepting the packet. This algorithm uses the following two parameters (r and b) for the traffic contract and one variable (TBC) for the internal usage.

- r: token rate, (corresponds to the monitored Maximum bitrate/Guaranteed bitrate).

- b: bucket size, (the upper bound of TBC, corresponds to bounded burst size).

- TBC(Token bucket counter): the number of given/remained tokens at any time.

In words, conformance according to a token bucket can be defined as: "Data is conformant if the amount of data submitted during any arbitrarily chosen time period T does not exceed (b+rT)".

The algorithm is described in the following.

Token bucket counter (TBC) is usually increased by "r" in each small time unit. However, TBC has upper bound "b" and the value of TBC shall never exceed "b".

When a packet #i with length L_i arrives, the receiver checks the current TBC. If the TBC value is equal to or larger than L_i, the packet arrival is judged compliant, i.e., the traffic is conformant. At this moment tokens corresponding to the packet length is consumed, and TBC value decreases by L_i.

When a packet #j with length L_j arrives, if TBC is less than L_j, the packet arrival is non-compliant, i.e., the traffic is not conformant. In this case, the value of TBC is not updated.
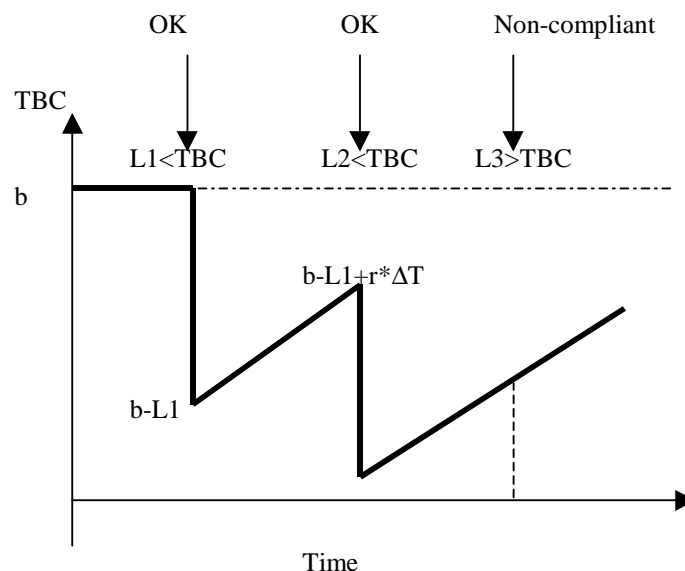

Figure B.1: Operation example of the reference conformance algorithm

Annex C (normative):
Determine which QoS profile is of highest QoS

In handovers from Release 1999 to GPRS Release 97/98 networks, it will be necessary to determine which PDP context of a set of PDP contexts provides the highest QoS, since, of a set of PDP contexts with the same APN and PDP address, all PDP contexts except the one with the highest QoS profile will be deactivated.

To determine which PDP context that has the highest QoS table C.1 is used. Only the PDP context(s) with the highest QoS ranking will be maintained and the rest will be deactivated. In a second step, if more than one PDP context remain, the PDP context with the highest value for the maximum bitrate attributes for downlink or uplink is selected. All PDP contexts except the PDP context(s) with the highest Maximum bitrate selected will be deactivated.

If more than one PDP context remain after the second step, all PDP contexts except that with the lowest NSAPI are deactivated.

Table C.1

QoS ranking	Traffic class	Traffic handling priority
1	Interactive	1
2	conversational	Not applicable
3	streaming	Not applicable
4	Interactive	2
5	Interactive	3
6	Background	Not applicable

Annex D (normative):
Determine Traffic Class weights in HLR QoS profile

The QoS profile in the subscription record represents the maximum QoS per PDP context to the associated APN. Subsequently, it shall be possible to negotiate all QoS parameters, including an appropriate Traffic Class for each QoS flow. This is valid for the first PDP context that is established as well as subsequent PDP contexts, i.e. this includes primary and secondary PDP contexts activations. The traffic classes have increasing weight according to the order background, interactive, streaming and conversational.

Annex E (informative): Change history

Change history								
Date	TSG #	TSG Doc.	CR	Rev	Cat	Subject/Comment	Old	New
12/2003	SP-22	SP-030654	0145	-	F	Radio Access Bearer Service Attributes for GERAN	5.10.0	6.0.0
03/2004	SP-23	SP-040033	0149	2	A	Correction to the use of delivery order set to yes	6.0.0	6.1.0
03/2004	SP-23	SP-040033	0150	2	F	ARP Clarification	6.0.0	6.1.0
03/2004	SP-23	SP-040033	0151	1	F	Removal of reliability class 1	6.0.0	6.1.0
12/2004	SP-26	SP-040912	0153	3	A	Clarification on delivery order set to no	6.1.0	6.2.0
06/2005	SP-28	SP-050334	0154	2	F	RAB Allocation/Retention Priority	6.2.0	6.3.0
06/2005	SP-28	SP-050334	0156	1	F	Addition of GERAN to the scope section	6.2.0	6.3.0
03/2006	SP-31	SP-060138	0159	-	F	Peak throughput class reference	6.3.0	6.4.0
06/2007	SP-36	-	-	-	-	Update to Rel-7 version (MCC)	6.4.0	7.0.0
09/2007	SP-37	SP-070539	0164	1	F	Support of higher maximum bitrate and guaranteed bit rate	7.0.0	7.1.0
12/2008	SP-42	-	-	-	-	Update to Rel-8 version (MCC)	7.1.0	8.0.0
12/2009	SP-46	-	-	-	-	Update to Rel-9 version (MCC)	8.0.0	9.0.0
06/2010	SP-48	SP-100327	0165	3	F	Addition of new QoS parameters evolved ARP and APN-AMBR	9.0.0	9.1.0
2011-03	SP-51	-	-	-	-	Update to Rel-10 version (MCC)	9.1.0	10.0.0
2011-06	SP-52	SP-110322	0168	1	A	Clarification of SRVCC usage of QoS	10.0.0	10.1.0

History

Document history		
V10.0.0	March 2011	Publication
V10.1.0	June 2011	Publication