

3GPP - TSG SA #5
11-13 October, 1999
Kyongju, Korea

Tdoc SP-99387

Title: 3G TS 23.107 V 2.0.0:
QoS Concept and Architecture
Date: 1999-10-06
Source: S2
Purpose: For information
Agenda Point: 5.2.3

The attached document contains version 2.0.0 of the 3G TS 23.107: *QoS Concept and Architecture*.

A earlier version of this document was presented as Technical Report 23.907 v.1.0.0 at SA#4. However, SA2 proposes to change it to Technical Specification at the same time it is proposed for approbation.

3G TS 23.107 2.0.0 (1999-10)

Technical Specification

3rd Generation Partnership Project; Technical Specification Group Services and System Aspects QoS Concept and Architecture (3G TS 23.107 version 2.0.0)



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

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

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

Reference

DTS/TSGS-0223107U

Keywords

UMTS, QoS

3GPP

Postal address

3GPP support office address

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

Internet

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

Contents

Contents	3
Intellectual Property Rights	5
Foreword.....	5
Introduction	Error! Bookmark not defined.
1 Scope	6
2 References	6
2.1 Normative References	6
3 Definitions, symbols and abbreviations	6
3.1 Definitions	6
3.2 Abbreviations	6
4 High Level Requirements.....	8
4.1 End User QoS Requirements	8
4.2 General Requirements for QoS	8
4.3 Technical Requirements for QoS.....	9
5 CS QoS in release '99.....	9
6 QoS Architecture	10
6.1 Overview of Different Levels of QoS.....	10
6.1.1 The End-to-End Service and UMTS Bearer Service	11
6.1.2 The Radio Access Bearer Service and the Core Network Bearer Service.....	12
6.1.3 The Radio Bearer Service and the Iu Bearer Service.....	12
6.2 QoS Management Functions in the Network	12
6.2.1 Description of functions	13
6.2.1.1 QoS management functions for UMTS bearer service in the control plane	13
6.2.1.2 Functions for UMTS bearer service in the user plane	13
6.2.2 Allocation of QoS management functions.....	14
6.2.2.1 QoS management functions for UMTS bearer service in the control plane	14
6.2.2.2 QoS management functions for the UMTS bearer service in the user plane	15
6.3 UMTS QoS Classes.....	16
6.3.1 Conversational class.....	16
6.3.2 Streaming class	17
6.3.3 Interactive class	17
6.3.4 Background class	17
6.4 QoS Parameters	18
6.4.1 Asymmetric Bearers	18
6.4.2 Sources of UMTS Bearer Service Parameters	18
6.4.3 UMTS Bearer Service Attributes.....	19
6.4.3.1 List of attributes	19
6.4.3.2 Attributes discussed per class	21
6.4.3.3 UMTS bearer attributes: summary	22
6.4.4 Radio Access Bearer Service Attributes.....	23
6.4.4.1 List of attributes	23
6.4.4.2 Attributes discussed per class	26
6.4.4.3 Radio Access Bearer attributes: summary	27
6.4.5 Radio Bearer Service Parameters	28
6.4.6 Iu Bearer Service Parameters	28
6.4.7 Core Network Bearer Service Parameters	28
6.5 Parameter Value Ranges.....	28
6.5.1 Ranges of UMTS Bearer Service Attributes	28
6.5.2 Ranges of Radio Access Bearer Service Attributes	29

7	Support of QoS Requirement During Inter-SGSN RA Update.....	30
8	QoS Parameter Mapping	31
8.1	From Application Parameters to UMTS Bearer Service Parameters	31
8.2	From UMTS Bearer Service Parameters to Radio Access Bearer Service Parameters	31
8.3	From UMTS Bearer Service Parameters to CN Bearer Service Parameters	32
9	Interworking	32
9.1	UMTS-GSM CS/GPRS	32
9.1.1	UMTS-GSM CS	32
9.1.2	UMTS-GPRS	32
9.2	UMTS-PSTN	33
9.3	UMTS-ISDN	33
9.4	UMTS-Internet	34
A. Annex:	Error resilience in real-time packet multimedia payloads	35
A.1.1	Factors affecting error resilience	35
A.2	Example figures	36
B. Annex:	Reference Algorithm for Conformance Definition of Bitrate.....	37
History.....		38

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 **3GPP members and non-members**, and can be found in SR 000 314: "*Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to 3GPP in respect of 3GPP standards*", which is available **free of charge** from the 3GPP Secretariat. Latest updates are available on the 3GPP Web server (<http://www.3gpp.org>).

Pursuant to the 3GPP IPR Policy, no investigation, including IPR searches, has been carried out by 3GPP. No guarantee can be given as to the existence of other IPRs not referenced in SR 000 314 (or the updates on the 3GPP Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Technical Specification has been produced by the 3GPP.

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 this TS, 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 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 specification;

1 Scope

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

2 References

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

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

2.1 Normative References

- [1] UMTS 23.110 version 3.1.0, Universal Mobile Telecommunications System (UMTS), UMTS Access Stratum - Services and Functions
- [2] UMTS 22.100 version 3.2.1, Universal Mobile Telecommunications System (UMTS), Service aspects, Service principles
- [3] UMTS 23.121 version 3.0.0, Evolution of the GSM platform towards UMTS
- [4] GSM 04.08 version 6.3.0, Digital cellular telecommunications system (Phase 2+), Mobile radio interface layer 3 specification
- [5] UMTS 22.105

3 Definitions, symbols and abbreviations

3.1 Definitions

For the purpose of this document the following definitions apply.

3.2 Abbreviations

For the purpose of this document the following abbreviations apply.

3G	3rd Generation
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
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
MS	Mobile Station
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 parameters has to be as small as possible.
- Derivation/definition of QoS attributes from the application requirements has to be simple.
- QoS attributes must be able to support all applications that are used, a certain number of applications have the characteristic of asymmetric nature between two directions, uplink/downlink.
- QoS definitions have to be future proof.
- QoS has to be provided end-to-end.

4.2 General Requirements for QoS

- QoS parameters (or mapping of them) should not be restricted to one or few external QoS control mechanisms but the QoS concept should be capable of providing different levels of QoS by using UMTS specific control mechanisms (not related to QoS mechanisms in the external networks).
- All parameters have to have unambiguous meaning.
- QoS mechanism have to allow efficient use of radio capacity.
- Allow independent evolution of Core and Access networks

- Allow evolution of UMTS network, (i.e., eliminate or minimise the impact of evolution of transport technologies in the wireline world)
- All parameter combinations have to have unambiguous meaning.

4.3 Technical Requirements for QoS

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

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

5 CS QoS in release '99

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, HLC) parameters
 - among the non-transparent data, **facsimile** is identified by the ITC
- c) **data, transparent**: from the ITC and CE parameters

For each of the above services an appropriate UMTS Bearer service shall be defined. The definition shall include exact UMTS bearer attribute values or list of supported values.

Note: This service mapping is the task of TSG N3.

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

Note: The mapping from GSM CC to UMTS RAB parameters 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.

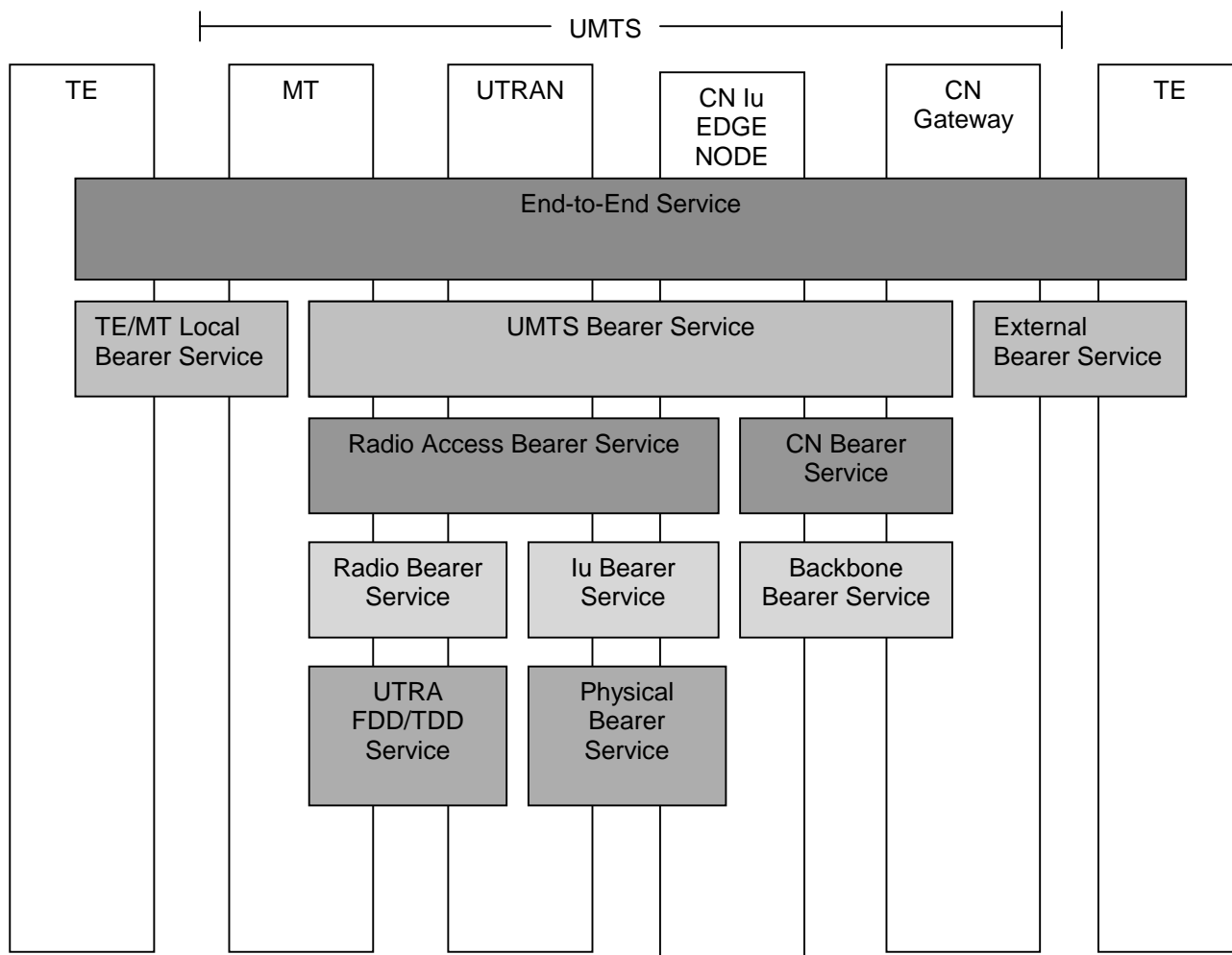


Figure 1: UMTS QoS Architecture

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 this document.

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

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

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

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

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

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

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

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, must 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 UTRAN using standardised attributes (see section 6.4.3).

In release '99, 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 Iu Edge Node with the CN Gateway to the external network. The role of this service is to efficiently control and utilise the backbone network in order to provide the contracted UMTS bearer service. The UMTS packet core network shall support different backbone bearer services for variety of QoS.

6.1.3 The Radio Bearer Service and the Iu Bearer Service

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

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

To support unequal error protection, UTRAN and MT must 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 Iu-Bearer Service together with the Physical Bearer Service provides the transport between UTRAN and CN. Iu bearer services for packet traffic shall provide different bearer services for variety of QoS.

6.1.4 The Backbone Network Service

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

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

6.2 QoS Management Functions in the Network

The purpose of this chapter is to give a comprehensive overview of functionality needed to establish, modify and maintain a UMTS Bearer Service with a specific QoS. The QoS management functions need not necessarily all to be standardised. Their allocation to the UMTS entities shall indicate the requirement for the specific entity to enforce the QoS commitments negotiated for the UMTS bearer service. The specific

realisation of these functions is implementation dependent and has only to maintain the specified QoS characteristics. The QoS management functions of all UMTS entities together shall ensure the provision of the negotiated service between the access points of the UMTS bearer service. The end-to-end service is provided by translation/mapping with UMTS external services.

6.2.1 Description of functions

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

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

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

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

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

6.2.1.2 Functions for UMTS bearer service in the user plane

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

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

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

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

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

unit traffic according to the QoS of the service. 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.

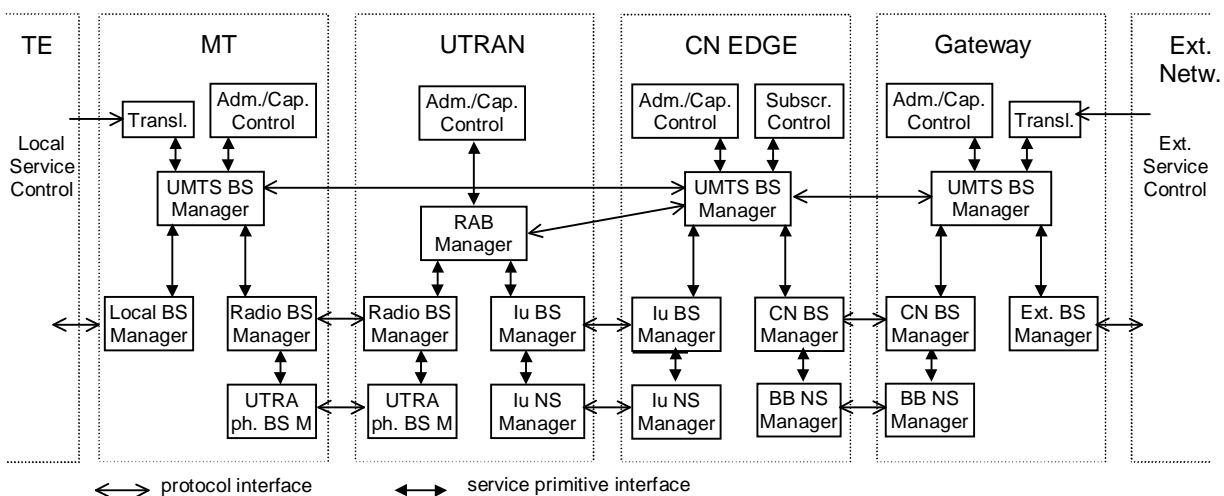


Figure 2: QoS management functions for UMTS bearer service in the control plane

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 Iu 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 Iu BS manager, its CN BS manager and the RAB manager in the UTRAN to provide the required services.

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

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

bearer service attributes into the external bearer service attributes and requests this service from the external BS manager.

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

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

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

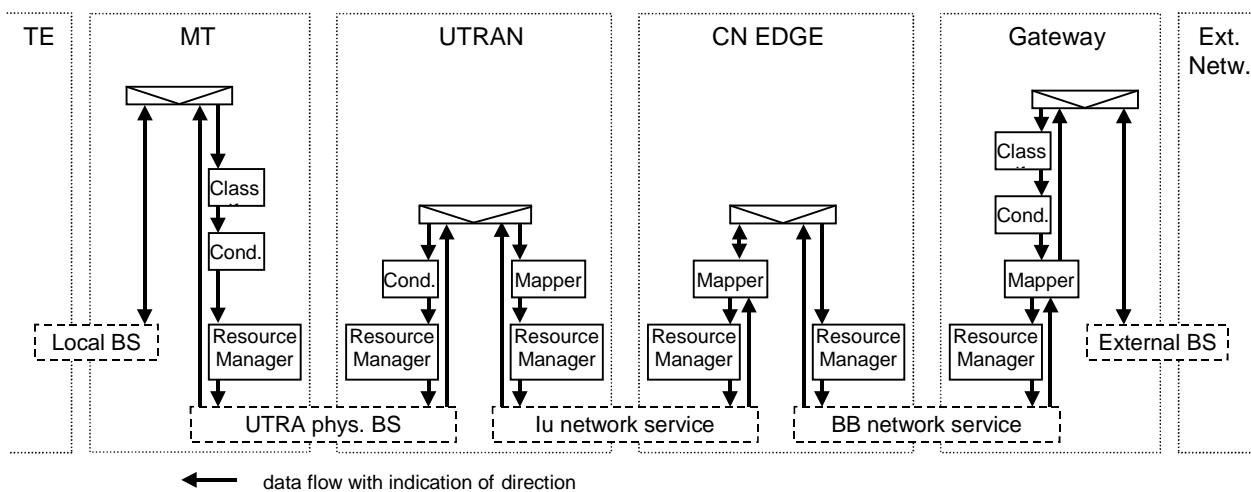


Figure 3: QoS management functions for the UMTS bearer service in the user plane

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

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

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

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

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

6.3 UMTS QoS Classes

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

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

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

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

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

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

6.3.1 Conversational class

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

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

Real time conversation - fundamental characteristics for QoS:

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

6.3.2 Streaming class

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

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

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

Real time streams - fundamental characteristics for QoS:

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

6.3.3 Interactive class

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

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

Interactive traffic - fundamental characteristics for QoS:

- request response pattern
- preserve payload content

6.3.4 Background class

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

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

Background traffic - fundamental characteristics for QoS:

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

Table 1: UMTS QoS classes

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

6.4 QoS Parameters

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

6.4.1 Asymmetric Bearers

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

6.4.2 Sources of UMTS Bearer Service Parameters

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

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

6.4.3 UMTS Bearer Service Attributes

6.4.3.1 List of attributes

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

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

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

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

Maximum bitrate [kbps]

Definition: maximum number of bits delivered by UMTS at a SAP within a 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.

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

Guaranteed bitrate [kbps]

Definition: guaranteed number of bits delivered by UMTS at a SAP within a 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 $k \cdot \text{Maximum SDU size}$. For release 99, $k=1$. A value of k greater than one Maximum SDU size may be specified in future releases to capture burstiness of sources. Signalling to specify the value of k may be provided in future releases.

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

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

Delivery order [y/n]

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

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

Maximum SDU size [bits]

Definition: the maximum allowed SDU size

[Purpose: The maximum SDU size is used for admission control and policing.]

SDU format information [bits]

Definition: list of possible exact sizes of SDUs

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

SDU error ratio

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

Note that 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 retransmission protocol on layer 2 and the error detection coding on layer 1.]

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 channel coding and error detection coding on layer 1.]

Delivery of erroneous SDUs (y/n/-)

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

Note: '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 frames with failed CRC on layer 1 shall be forwarded or not.]

Transfer delay [s]

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

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

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

Traffic handling priority

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

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

Allocation/Retention Priority

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

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

6.4.3.2 Attributes discussed per class

Conversational class

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

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

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

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 UTRAN without RLC re-transmissions. Hence, UTRAN transport is more efficient and thereby cheaper if RLC PDU size is adapted to UMTS bearer SDU size (RLC transparent mode). This motivates the use of **SDU 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 must 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. 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 must 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 UTRAN/CN. Instead, **traffic handling priority** is used. SDUs of a UMTS bearer with higher traffic handling priority is given priority over SDUs of other bearers within the interactive class, through UMTS-internal scheduling.

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

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

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 class are summarised. Observe that traffic class is an attribute itself.

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

Traffic class	Conversational class	Streaming class	Interactive class	Background class
Maximum bitrate	X	X	X	X
Delivery order	X	X	X	X
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

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

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

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

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

[Purpose: By including the traffic class itself as an attribute, UTRAN 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 UTRAN 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.

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

Guaranteed bitrate [kbps]

Definition: guaranteed number of bits delivered at a SAP within a 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 k Maximum SDU size. For Release 99, k = 1. A value of k greater than one Maximum SDU size may be specified in future releases to capture burstiness of sources. Signalling to specify the value of k may be provided in future releases.

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

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

Delivery order [y/n]

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

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

Maximum SDU size [bits]

Definition: the maximum allowed SDU size

[Purpose: The maximum SDU size is used for admission control and policing.]

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.

Note: SDU format information is used by UTRAN 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: UTRAN 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, UTRAN needs to know the exact format of SDU payload to be able to demultiplex the SDU onto different radio bearer services.]

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 that 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 retransmission protocol on layer 2 and the error detection coding on layer 1]

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 channel coding and error detection coding on layer 1. 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: '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 frames with failed CRC on layer 1 shall be forwarded or discarded.]*

Transfer delay [s]

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

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

Traffic handling priority

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

[Purpose: Within the interactive class, there is a definite need to differentiate between bearer qualities. This is handled by using the traffic handling priority attribute, to allow UTRAN to

schedule traffic accordingly. By definition, priority is an alternative to absolute guarantees, and thus these two attribute types cannot be used together for a single bearer.]

Allocation/Retention Priority

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

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

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, UTRAN may, based on experience, calculate a statistical multiplex gain for use in admission control on the radio and lu interfaces.]

6.4.4.2 Attributes discussed per class

Conversational class

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

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, must 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.

Streaming class

If the RAB carries streaming speech, **Source statistics descriptor** can be set, which allows UTRAN to calculate a statistical multiplexing gain on radio and lu 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, must 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.

Other classes

The RAB attribute sets and their use in, interactive and background classes are identical to those of UMTS bearer services (Section 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 class are summarised. Observe that traffic class is an attribute itself.

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

Traffic class	Conversational class	Streaming class	Interactive class	Background class
Maximum bitrate	X	X	X	X
Delivery order	X	X	X	X
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		

6.4.5 Radio Bearer Service Parameters

Note: Defining the radio bearer service parameters is a task for RAN WG2.

6.4.6 Iu Bearer Service Parameters

The Iu-Bearer Service together with the Physical Bearer Service provides the transport between UTRAN and CN. Iu bearer services for packet traffic shall provide different 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 IP based Iu 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 Parameters

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 Parameter Value Ranges

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.

Traffic class	Conversational class	Streaming class	Interactive class	Background class
Maximum bitrate [kbps]	<2000 (1) (2)	<2000 (1) (2)	< 2000 - overhead (2) (3)	<2000 - overhead (2) (3)
Delivery order	Yes/No	Yes/No	Yes/No	Yes/No
Maximum SDU size [octets]	<1500 (4)	<1500 (4)	<1500 (4)	<1500 (4)
SDU format information	(5)	(5)		
Delivery of erroneous SDUs	Yes/No/- (6)	Yes/No/- (6)	Yes/No/- (6)	Yes/No/- (6)
Residual BER	$5 \cdot 10^{-2}$, 10^{-2} , 10^{-3} , 10^{-4} (7)	$5 \cdot 10^{-2}$, 10^{-2} , 10^{-3} , 10^{-4} , 10^{-5} , 10^{-6} (7)	$4 \cdot 10^{-3}$, 10^{-5} , $6 \cdot 10^{-8}$ (8) (7)	$4 \cdot 10^{-3}$, 10^{-5} , $6 \cdot 10^{-8}$ (8) (7)

SDU error ratio	$10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$ (7)	$10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$ (7)	$10^{-3}, 10^{-4}, 10^{-6}$ (7)	$10^{-3}, 10^{-4}, 10^{-6}$ (7)
Transfer delay [ms]	100 – maximum value(7)	500 – maximum value (7)		
Guaranteed bit rate [kbps]	<2000 (1) (2)	<2000 (1) (2)		
Traffic handling priority			1,2,3 (9)	
Allocation/Retention priority	1,2,3 (9)	1,2,3 (9)	1,2,3 (9)	1,2,3 (9)

- 1) Bitrate of 2000 kbps requires that UTRAN operates in transparent RLC protocol mode, in this case the overhead from layer 2 protocols is negligible.
- 2) The granularity of the bit rate parameters must be studied. Although the UMTS network has capability to support a large number of different bitrate values, the number of possible values must be limited not to unnecessarily increase the complexity of for example terminals, charging and interworking functions. Exact list of supported values shall be defined together with S1, N1, N3 and R2.
- 3) Impact from layer 2 protocols on maximum bitrate in non-transparent RLC protocol mode shall be estimated.
- 4) Maximum SDU size shall at least allow UMTS network to support external PDUs having as high MTU as Internet/Ethernet (1500 octets). The need for higher values must be investigated by N1, N3, S1, R2, R3.
- 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) 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 indicative. Exact values on Residual BER, SDU error ratio and transfer delay shall defined together with S1, N1, N3 and R2.
- 8) Values are derived from CRC lengths of 8, 16 and 24 bits on layer 1.
- 9) Number of priority levels shall be further analysed by S1, N1 and N3.

6.5.2 Ranges of Radio Access Bearer Service Attributes

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

Traffic class	Conversational class	Streaming class	Interactive class	Background class
Maximum bitrate [kbps]	<2000 (1) (2)	<2000 (1) (2)	< 2000 - overhead (2) (3)	<2000 - overhead (2) (3)
Delivery order	Yes/No	Yes/No	Yes/No	Yes/No
Maximum SDU size [octets]	<1500 (4)	<1500 (4)	<1500 (4)	<1500 (4)
SDU format information	(5)	(5)		
Delivery of erroneous SDUs	Yes/No/-	Yes/No/-	Yes/No/-	Yes/No/-

Residual BER	$5 \cdot 10^{-2}$, 10^{-2} , 10^{-3} , 10^{-4} (6)	$5 \cdot 10^{-2}$, 10^{-2} , 10^{-3} , 10^{-4} , 10^{-5} , 10^{-6} (6)	$4 \cdot 10^{-3}$, 10^{-5} , $6 \cdot 10^{-8}$ (6) (7)	$4 \cdot 10^{-3}$, 10^{-5} , $6 \cdot 10^{-8}$ (6) (7)
SDU error ratio	10^{-2} , 10^{-3} , 10^{-4} , 10^{-5} (6)	10^{-2} , 10^{-3} , 10^{-4} , 10^{-5} (6)	10^{-3} , 10^{-4} , 10^{-6} (6)	10^{-3} , 10^{-4} , 10^{-6} (6)
Transfer delay [ms]	80 – maximum value(6)	500 – maximum value (6)		
Guaranteed bit rate [kbps]	<2000 (1) (2)	<2000 (1) (2)		
Traffic handling priority			1,2,3 (8)	
Allocation/Retention priority	1,2,3 (8)	1,2,3 (8)	1,2,3 (8)	1,2,3 (8)
Source statistic descriptor	Speech/unknown	Speech/unknown	Speech/unknown	Speech/unknown

- 1) Bitrate of 2000 kbps requires that UTRAN operates in transparent RLC protocol mode, in this case the overhead from layer 2 protocols is negligible.
- 2) The granularity of the bit rate parameters must be studied. Although the UMTS network has capability to support a large number of different bitrate values, the number of possible values must be limited not to unnecessarily increase the complexity of for example terminals, charging and interworking functions. Exact list of supported values shall be defined together with S1, N1, N3 and R2.
- 3) Impact from layer 2 protocols on maximum bitrate in non-transparent RLC protocol mode shall be estimated.
- 4) Maximum SDU size shall at least allow UMTS network to support external PDUs having as high MTU as Internet/Ethernet (1500 octets). The need for higher values must be investigated by N1, N3, S1, R2, R3.
- 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 indicative. Exact values on Residual BER, SDU error ratio and transfer delay shall defined together with S1, N1, N3 and R2.
- 7) Values are derived from CRC lengths of 8, 16 and 24 bits on layer 1.
- 8) Number of priority levels shall be further analysed by S1, N1 and N3.

7 Support of QoS Requirement During Inter-SGSN RA Update

Support of QoS within the GSM/GPRS R99 network will require enhancements to the standard to enable QoS requirements to be maintained when mobility of the UE requires a change of serving elements within the network (e.g. change of BTS/BSC or nodeB/RNS and SGSN). These are currently not supported in GSM/GPRS R99 and therefore the next release shall support these enhancements.

QoS requirements (especially those for classes such as Conversational and Streaming) shall be maintained during inter SGSN RA Update within the UMTS network (note: Pre R99 inter SGSN RA update mechanisms involving C7 signalling to the HLR, GGSN would need enhancement to satisfy some QoS classes identified).

8 QoS Parameter Mapping

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

8.1 From Application Parameters to UMTS Bearer Service Parameters

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

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

When establishing a UMTS bearer and the underlying Radio Access Bearer for support of a service request, some parameter on UMTS level does typically not have the same value as corresponding parameter on Radio Access Bearer level. For example requested transfer delay for the UMTS bearer must typically be larger than 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 parameters/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: 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
- Allocation/Retention priority
- Maximum SDU size
- SDU format information
- NOTE: 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 parameters 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 implementational and depends for example on network dimensioning.

- **Residual BER** for Radio Access Bearer service must be reduced with the bit errors introduced in the core network, by Core Network Bearer service.
- **SDU error ratio** for Radio Access Bearer service must be reduced with the errors introduced in the core network, by Core Network Bearer service.
- **Transfer delay** for Radio Access Bearer service must be reduced with the delay introduced in the core network, e.g. on transmission links or in a codec resident in the Core Network.

The following parameters/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 Parameters to CN Bearer Service Parameters

Note: This is operator's choice.

9 Interworking

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

9.1 UMTS-GSM CS/GPRS

9.1.1 UMTS-GSM CS

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

9.1.2 UMTS-GPRS

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

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

Conversational Class

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

An appropriate use of GPRS parameters:

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

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

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

Streaming Class

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

An appropriate use of GPRS parameters:

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

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

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

Interactive Class

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

An appropriate use of GPRS parameters:

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

Background Class

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

An appropriate use of GPRS parameters:

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

9.2 UMTS-PSTN

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

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

9.3 UMTS-ISDN

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

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

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

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

A. Annex: 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 4.



Figure 4: Structure of H.323 packet.

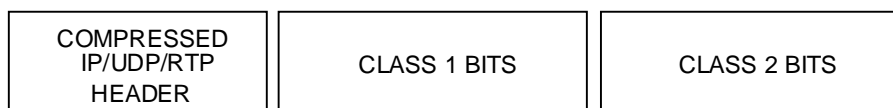


Figure 5: Structure of compressed H.323 packet.
Class 1 bits can tolerate medium BER; Class 2 bits can tolerate high BER.

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 5. 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 parameters 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 must be transmitted reliably (i.e. residual bit errors should result in a lost packet).

B. Annex: 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 must 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.

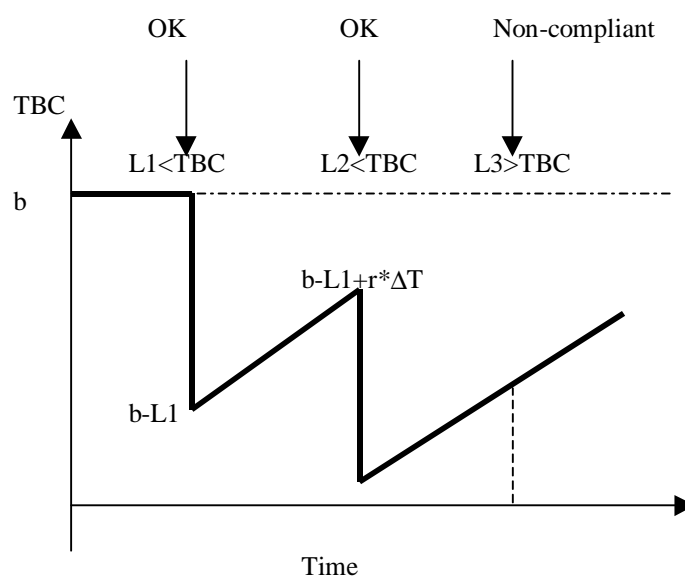


Figure 6. Operation example of the reference conformance algorithm.

History

Document History		
October 1998		Scope agreed
November 1998	0.1.0	ToR agreed
January 1999	0.2.0	Parameters, traffic classes and management functions added
April 1999	0.3.0	Moved to template
April 1999	0.4.0	Parameter definitions and sources added and annexes removed
May 1999	0.5.0	RAB attributes added
May 1999	1.0.0	Version approved to be changed to 1.0.0 by TSG SA2 #5
June 1999	1.1.0	QoS management functions updated
July 1999	1.2.0	UMTS and radio access bearer parameters updated, annex added
July 1999	1.3.0	UMTS QoS user plane management functions figure updated and CS domain QoS mechanisms have been added
September 1999	1.4.0	Reliability attribute changed, bitrate definitions updated, Diffserv usage clarified, SDU periodicity clarified
September 1999	1.5.0	Cleanup moving and removing chapters, Clarifications added to several places and text updated.
September 1999	1.6.0	Few editorial mistakes corrected, one mistake concerning UEP corrected.
October 1999	1.6.1 and 1.6.2	A few typos and editorial errors corrected.
October 1999	TS 23.107 v.2.0.0	Changed from TR to TS. Technical content identical to 23.907 v.1.6.2.