Technical Specification Group Services and System Aspects
Meeting #10, Bangkok, Thailand, 11-14 December 2000

TSGS#10(00)0594

3GPP TSG-SA2 #15
November 13-17, 2000
Makuhari, Japan

S2-002077

Title: TS 23.207 Version 1.0.0 (End-to-End QoS Concept and Architecture)

To: TSG-SA

Source: TSG-SA-WG2

For: Information

*Contact: (Editor) Johnson Oyama
Johnson.Oyama@era.ericsson.se

TS 23.207 Version 1.0.0 (Title: End-to-End QoS Concept and Architecture) is submitted to TSG-SA for information.

3G TS 23.207 V1.0.0 (2000-11)

Technical Specification

3rd Generation Partnership Project;
Technical Specification Group Services and System Aspects;
End-to-End QoS Concept and Architecture

Keywords

<keyword[, keyword]>

3GPP

Postal address

3GPP support office address

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

Internet

http://www.3gpp.org

09/21/00 7:00 PM

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 end-to-end Quality of Service involving GPRS and complements TS23.107 which describes the framework for Quality of Service within UMTS. The end-to-end QoS architecture is provided in Figure 1. The document describes the interaction between the TE/MT Local Bearer Service, the GPRS Bearer Service, and the External Bearer Service, and how these together provide Quality of Service for the End-to-End Service. The document also describes IP level mechanisms necessary in providing end-to-end Quality of Service involving GPRS networks, including possible interaction between the IP level and the GPRS level, as well as the application level and the IP level.

In contrast to the TS23.107, the present document is only applicable to GPRS packet switched access services, and includes aspects of interworking to the IM subsystem as well as PSTN and other networks. The document does not cover the circuit switched access services.

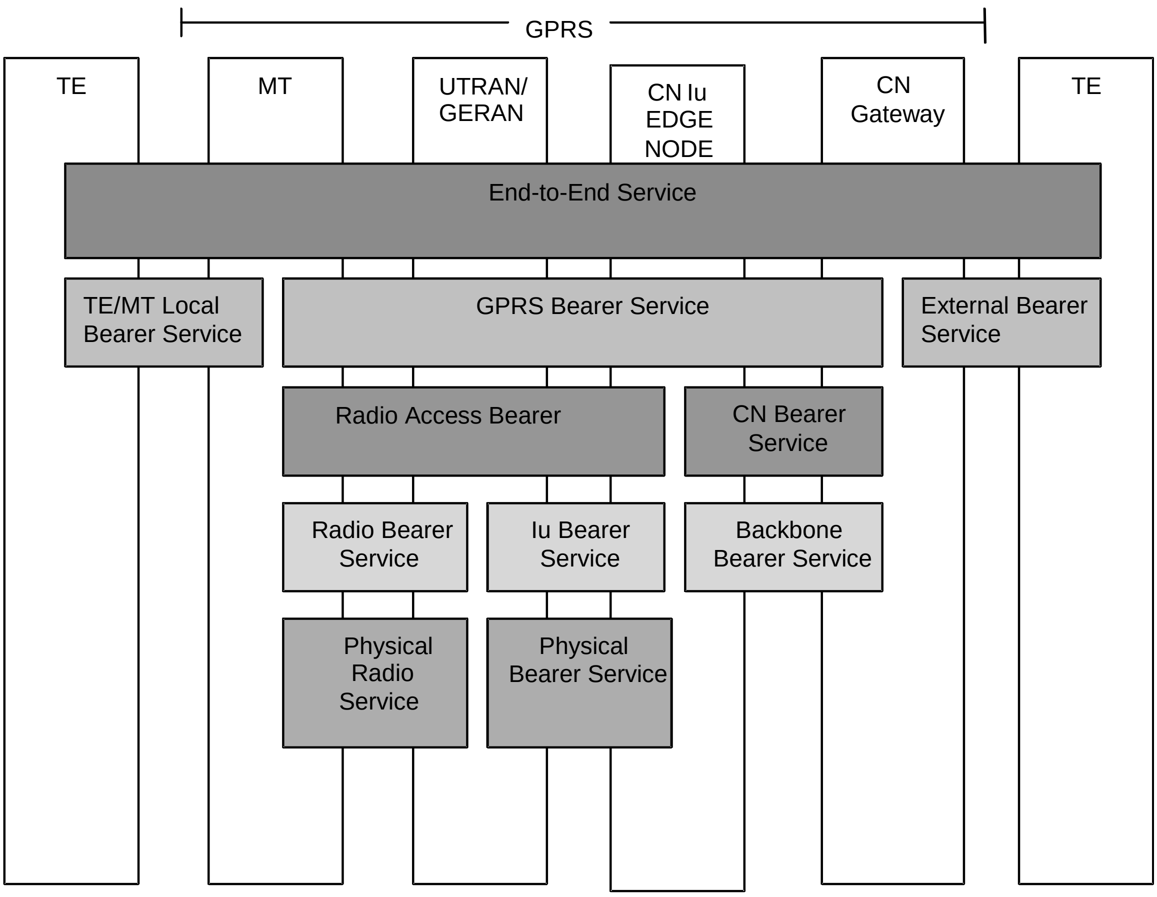

Figure 1: End-to-End QoS Architecture

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.

[1] xxx

3 Abbreviations

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

[Editorial Note: The following preliminary list of abbreviations are taken from the TS23.107, and need further verification regarding applicability to the current TS.]

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
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 for End-to-End IP QoS

4.1 End-to-End QoS Negotiation Requirements

- The UMTS QoS negotiation mechanisms used for providing end-to-end QoS shall be backward compatible with UMTS Release 99.
- The UMTS QoS negotiation mechanisms used for providing end-to-end QoS shall not make any assumptions about the situation in external networks which are not within the scope of 3GPP specifications.
- The UMTS QoS negotiation mechanisms used for providing end-to-end QoS shall not make any assumptions about application layer signalling protocols.
- No changes to non-UMTS specific QoS negotiation mechanisms.
- The UMTS QoS negotiation mechanisms used for providing end-to-end QoS shall not make any assumptions about applications which may be used on terminal equipment attached to mobile terminals.
- Unnecessary signalling complexity and processing complexity in the network elements as well as the mobile terminal shall be avoided.
- Unnecessary signalling traffic due to end-to-end QoS negotiation shall be avoided.
- Methods for user authentication as well as billing and charging mechanisms related to the end-to-end QoS negotiation shall be kept as simple as possible.
- Minimum changes to network architecture and mechanisms due to introduction of end-to-end QoS negotiation.
- The UMTS network shall be able to negotiate end-to-end QoS also for mobile terminals and applications which are not able to use QoS negotiation mechanisms other than the ones provided by UMTS.
- It shall be possible for an application on the external device to request end-to-end QoS

4.2 QoS Policy Requirements

- The UMTS policy mechanisms described in TS 23.060 shall be used for control of the UMTS bearers.
- Interaction between UMTS bearer services and IP bearer services shall only occur at the translation function in the UE and GGSN.

5 End-to-End IP QoS Architecture

5.1 QoS Management Functions in the Network

To provide IP QoS end-to-end, it is necessary to manage the QoS within each domain. An IP BS Manager is used to control the external IP bearer service. Due to the different techniques used within the IP network, this communicates to the UMTS BS manager through the Translation function.

To enable coordination between events in the application layer and resource management in the IP bearer layer, an element called IP Policy Control is used as a logical policy decision element. It is also possible to implement a policy decision element internal to the IP BS Manager in the GGSN. The IP policy architecture does not mandate the policy decision point to be external to the GGSN.

Whenever resources not owned or controlled by the UMTS network are required to provide QoS, it is necessary to interwork with the external network that controls those resources. Interworking may be realised in a number of ways, including:

- signalling along the flow path (e.g. RSVP, LDP).
- packet marking or labelling along the flow path (e.g. DiffServ, MPLS)
- interaction between Policy Control and/or Resource Management elements.
- Service Level Agreements enforced by the border routers between networks.

For the policy control the following guidelines should apply:

- The IP policy framework employed in UMTS should, as far as possible, conform to IETF "Internet Standards". The IETF policy framework may be used for policy decision, authorization, and control of the IP level functionality, at both user and network level.
- There should be separation between the scope and roles of the UMTS policy mechanisms and the IP policy framework. This is to facilitate separate evolution of these functions.

5.1.1 Description of functions

5.1.1.1 QoS management functions for end-to-end IP QoS in UMTS Network

NOTE: The end-to-end QoS management functions do not cover the cases of a circuit switched service, or an IP service interworking with an ATM service at the gateway node.

IP BS Manager uses standard IP mechanisms to manage the IP bearer service. These mechanisms may be different from mechanisms used within the UMTS, and may have different parameters controlling the service. The translation/mapping function provides the interworking between the mechanisms and parameters used within the UMTS bearer service and those used within the IP bearer service, and interacts with the IP BS Manager.

If an IP BS Manager exists both in the UE and the Gateway node, it is possible that these IP BS Managers communicate directly with each other by using relevant signalling protocols.

The required options in the table define the minimum functionality that shall be supported by the equipment in order to allow multiple network operators to provide interworking between their networks for end-to-end QoS. Use of the optional functions listed below, other mechanisms which are not listed (eg over-provisioning), or combinations of these mechanisms are not precluded from use between operators.

The IP BS Managers in the UE and GGSN provide the set of capabilities for the IP bearer level as shown in Table 1. Provision of the IP BS Manager is optional in the UE, and required in the GGSN.

Table 1: IP BS Manager capability in the UE and GGSN

Capability	UE	GGSN
DiffServ Edge Function	Optional	Required
RSVP/IntServ	Optional	Optional
IP Policy Enforcement Point	Optional	Required (*)

(*)Although the capability of IP policy enforcement is required within the GGSN, the control of IP policy through the GGSN is a network operator choice. Where the APN is not located at the GGSN, the location of policy enforcement point is for further investigation.

[Editorial note: There is a need to clarify the relationship between the APN and the access point to the newtork identified by the APN. The concept of access point has to be clarified.]

Figure 2 shows the scenario for control of an IP service using IP BS Managers in both possible locations in the UE and Gateway node. The figure also indicates the optional communication path between the IP BS Managers in the UE and the Gateway node.

IP Policy Control is a logical policy decision element which uses standard IP mechanisms to implement policy in the IP bearer layer. These mechanisms may be conformant to, for example, the framework defined in IETF [RFC2753] "A Framework for Policy-based Admission Control" where the IP Policy Control is effectively a Policy Decision Point (PDP). The IP Policy Control makes decisions in regard to network based IP policy using policy rules, and communicates these decisions to the IP BS Manager in the GGSN, which is the IP Policy Enforcement Point (PEP).

A protocol interface between the IP Policy Control and application servers/proxies (e.g. local SIP proxy) supports the transfer of policy related information from the application layer to the policy decision point.

[Editorial Note: The exact mechanisms, protocols whether proprietary or standardized, and how they are used are for further study.]

A protocol interface between the IP Policy Control and GGSN supports the transfer of information and policy decisions between the policy decision point and the IP BS Manager in the GGSN.

[Editorial Note: The exact mechanisms, protocols whether proprietary or standardized, and how they are used are for further study. One possible candidate is the COPS protocol [RFC2748] which describes a simple query and response protocol that can be used to exchange policy information between a policy server (PDP) and its client (PEP). Where RSVP is used as the signalling protocol in the IP bearer level, a COPS protocol variant carrying embedded RSVP information, i.e., COPS-RSVP, defined in [RFC2749] may be used.]

The IP Policy Control bases policy decisions only on information obtained from nodes / elements within its domain or from nodes with which it has a trust relationship. The IP Policy Control needs to be in the same domain as the GGSN or have a trust relationship with the GGSN.

NOTE: Currently in IETF, inter-domain policy interactions are not defined.

NOTE: The security issues regarding the trust relationship between the nodes / elements is outside the scope of this chapter.

[Editorial Note: Additionally, the IP Policy Control may have protocol interfaces to other devices (e.g., AAA, bandwidth broker) which support transfer of information (e.g., authentication, availability of resources, etc.) for use in policy decisions. These are for further study.]

[Editorial Note: Where the access point of the APN is not located at the GGSN, the location of policy enforcement point is for further investigation. The IP policy architecture for cases where the access point of the APN is located in a third party network, e.g., a corporate network, is for further study.]

5.1.1.2 Interaction to External Networks

Within the UMTS network, there is resource management performed by various nodes in the admission control decision. The resources considered here are under the direct control of the UMTS network.

In IP Networks, it is also necessary to perform resource management to ensure that resources required for a service are available. Where the resources for the IP Bearer Service to be managed are not owned by the UMTS network, the resource management of those resources would be performed through an interaction between the UMTS network and that external network.

In addition, where the UMTS network is also using external IP network resources as part of the UMTS bearer service (for example for the backbone bearer service), it may also be necessary to interwork with that network.

The GGSN shall support DiffServ edge functionality and be able to shape upstream traffic. There are a number of other mechanisms provided to support interoperator interworking, some of which are given in Annex D.

5.1.2 Allocation of QoS management functions

5.1.2.1 QoS management functions for end-to-end IP QoS

The QoS management functions for controlling the external IP bearer services and how they relate to the UMTS bearer service QoS management functions are shown in Figure 2.

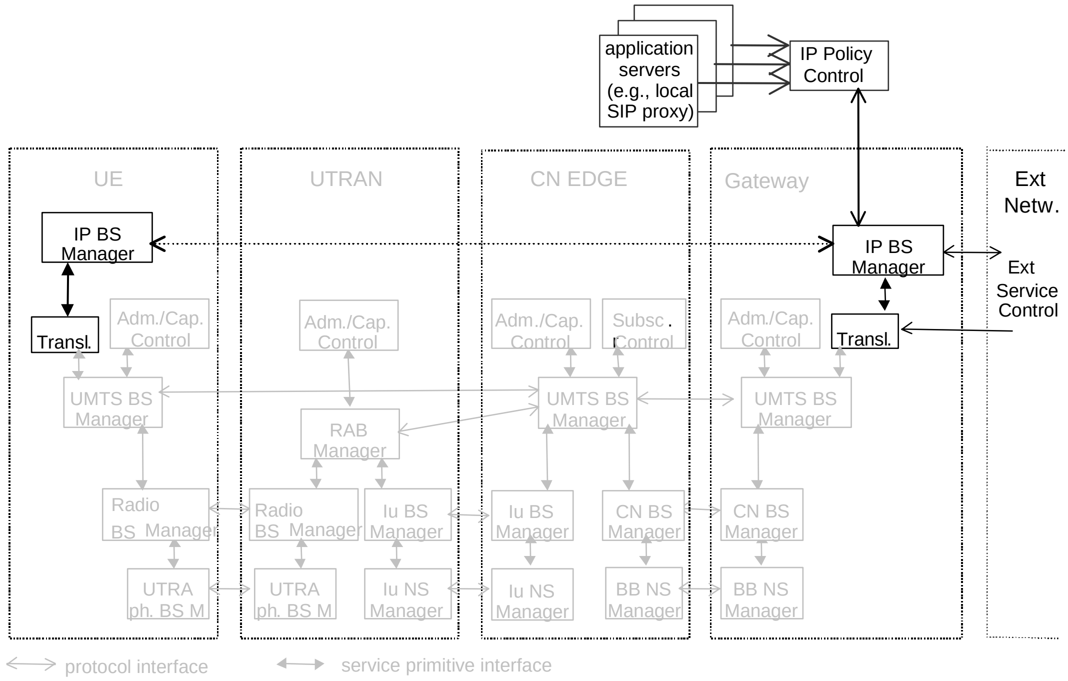

Figure 2: QoS management functions for UMTS bearer service in the control plane and QoS management functions for end-to-end IP QoS

NOTE: The dimmed boxes in Figure 2 are clarified in TS23.107.

[Editorial note: Figure 2 and this chapter shows UE only as a combined element. This TS also need to consider the case where the TE and MT are split. A section providing the split and the distribution of functionality need to be added to this TS and is for further study. Standardization of the interface between the TE and MT is the responsibility of the 3GPP working group TSG T2, and is outside the scope of this TS.]

[Editorial Note: Elements external to the nodes are used to highlight and explain possible solutions to requirements that have been identified. If elements or interfaces are specified or mandated within 3GPP, they shall be included in the Reference Architecture.]

5.2 QoS Parameters

6 QoS Parameter Mapping

7 End-to-End QoS Procedures

Annex A (Informative): QoS Conceptual Models

A.1 Introduction

There are many different end-to-end scenarios that may occur from a UE connected to a UTMS network. The following examples depict how end-to-end QoS will be delivered for a number of scenarios that are considered to be significant.

In all the scenarios presented below, the network architecture is as shown in Figure A.1 below.

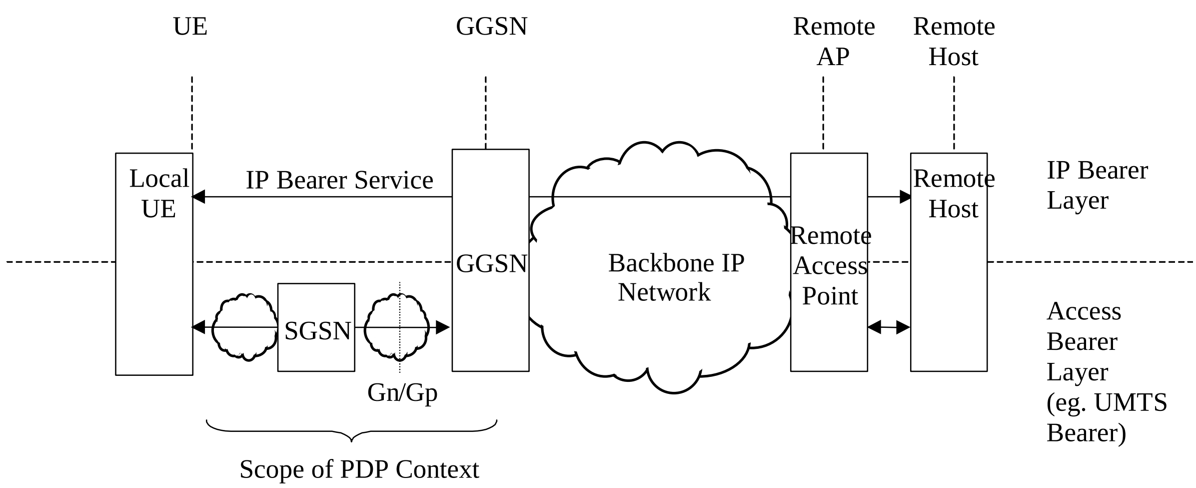

Figure A.1: Network Architecture for QoS Conceptual Models

Notes:

- Although the backbone IP network is shown as a single domain, it may consist of a number of separate domains.
- The structure of the Local UE is not specified. It includes cases from a simple host, to a gateway to a network such as a LAN. If the UE is acting as a gateway, it is responsible for providing the IP BS Management towards the extended network.
- The remote side is shown as a simple host. Other more complex cases on the remote side such as a private LAN with over-provisioning, or possibly LAN priority marking, and DiffServ and/or RSVP capable routing elements is not depicted. It is envisaged however that interworking between the QoS mechanisms in a more complex remote user side could also be performed with some similarities to the mechanisms shown at the local side.
- The GGSN and the APN are co-located in these scenarios.

The reference point shown at the UE is at the interface to the UE. Within the UE, the QoS control could be derived from any of the mechanisms that occur across that reference point, or it could use a different mechanism internally.

Although the scenarios currently identified are mainly using DiffServ in the backbone IP network (RSVP is indicated as an alternative in scenario 4), it is not mandated that DiffServ must be used in the backbone IP network. Other mechanisms, for example, over-provisioning and aggregated RSVP may be used.

A.2 Scenarios

[Editorial Note: the precedence and sequence of the different phases of session / bearer establishment need further study.]

NOTE: Scenario 5 and 6 is reserved for the IP multimedia services involving, e.g., SIP signalling , IP policy control, and subscription checking.

A.2.1 Scenario 1

The UE does not provide an IP BS Manager. The end-to-end IP QoS bearer service towards the remote terminal is controlled from the GGSN.

The scenario assumes that the GGSN supports DiffServ edge functions, and the backbone IP network is DiffServ enabled.

In this scenario, the control of the QoS over the UMTS access network (from the UE to the GGSN) may be performed either from the terminal using the PDP context signalling, or from the SGSN by subscription data.

The IP QoS for the downlink direction is controlled by the remote terminal up to the GGSN. The GGSN will apply receiver control DiffServ edge functions and can reclassify the data (remarking the DiffServ Code Point (DSCP)). This may affect the QoS applied to the data over the UMTS access (the TFT may use the DSCP to identify the data to be allocated to the PDP context).

The end-to-end QoS is provided by a local mechanism in the UE, the PDP context over the UMTS access network, DiffServ through the backbone IP network, and DiffServ in the remote access network in the scenario shown in the figure below. The GGSN provides the interworking between the PDP context and the DiffServ function. However, the interworking may use information about the PDP context which is established, or be controlled from static profiles, or dynamically through other means such as proprietary HTTP based mechanisms. The UE is expected to be responsible for the control of the PDP context, but this may instead be controlled from the SGSN by subscription.

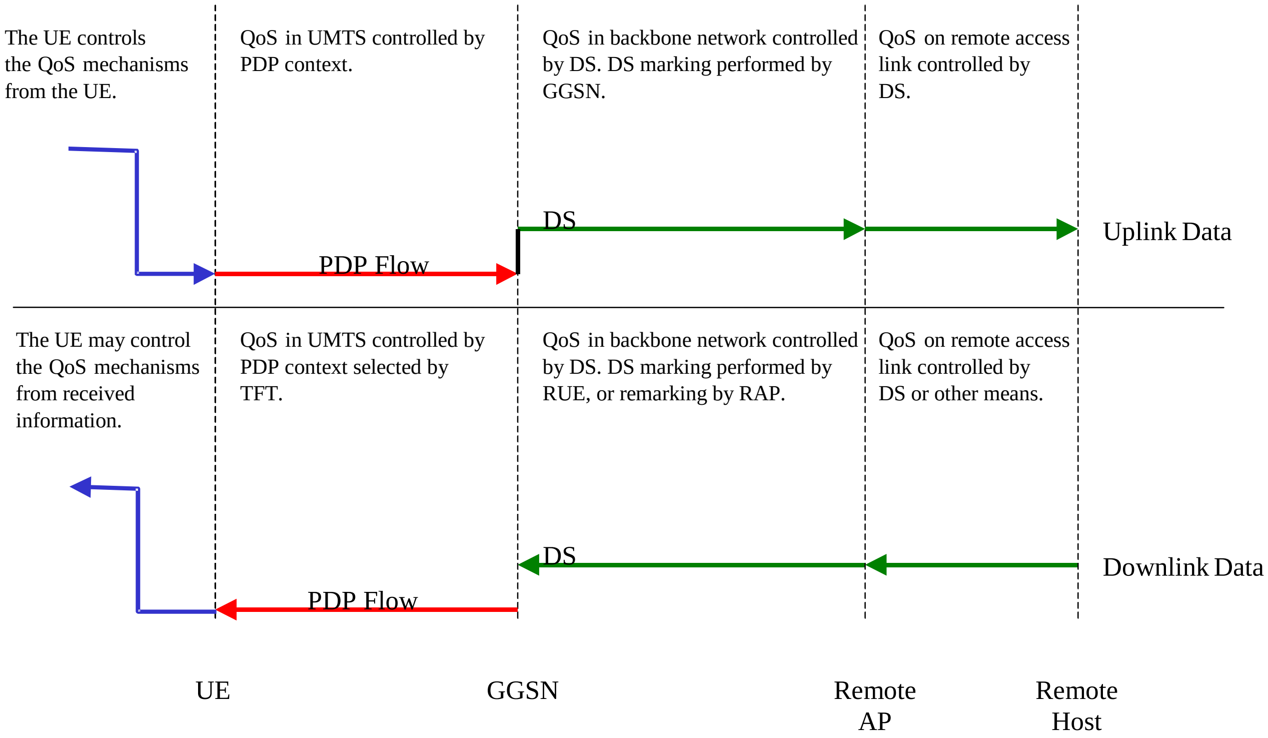

Figure A.2: Local UE does not provide IP BS Manager

Notes:

- The solid horizontal lines indicate the mechanism that is providing QoS for the flow of data in the direction indicated.
- The dashed horizontal lines indicate where QoS control information is passed that is not directly controlling the QoS in that link/domain.
- The arrows on the horizontal lines indicate nodes that receive information about QoS from that mechanism, even if that mechanism is not used to control the QoS over that link/domain.
- The solid vertical lines indicate interworking between the different mechanisms.

No solid vertical line is shown from DiffServ to PDP flow on the downlink at the GGSN. The TFT determines the QoS applicable over the UMTS access. However, the configuration of the TFT may use the DiffServ to select the PDP context to be applied, so there may be interworking between DiffServ and the PDP Flow via the TFT filters.

A.2.2 Scenario 2

The UE performs an IP BS function which enables end-to-end QoS without IP layer signalling towards the IP BS function in the GGSN, or the remote terminal.

The scenario assumes that the UE and GGSN support DiffServ edge functions, and that the backbone IP network is DiffServ enabled.

In this scenario, the control of the QoS over the UMTS access network (from the UE to the GGSN) may be performed either from the terminal using the PDP context signalling. Alternatively, subscription data accessed by the SGSN may override the QoS requested via signalling from the UE.

In this scenario, the terminal supports DiffServ to control the IP QoS through the backbone IP network.

The IP QoS for the downlink direction is controlled by the remote terminal up to the GGSN. The PDP context controls the QoS between the GGSN and the UE. The UE may apply DiffServ edge functions to provide the DiffServ receiver control. Otherwise, the DiffServ marking from the GGSN will determine the IP QoS applicable at the UE.

The end-to-end QoS is provided by a local mechanism in the UE, the PDP context over the UMTS access network, DiffServ through the backbone IP network, and DiffServ in the remote access network in the scenario shown in Figure A.3 below. The UE provides control of the DiffServ, and therefore determines the appropriate interworking between the PDP context and DiffServ.

The GGSN DiffServ edge function may overwrite the DSCP received from the UE, possibly using information regarding the PDP context which is signalled between the UMTS BS managers and provided through the translation/mapping function to the IP BS Manager.

Note that DiffServ control at the Remote Host is shown in this example. However, other mechanisms may be used at the remote end, as demonstrated in the other scenarios.

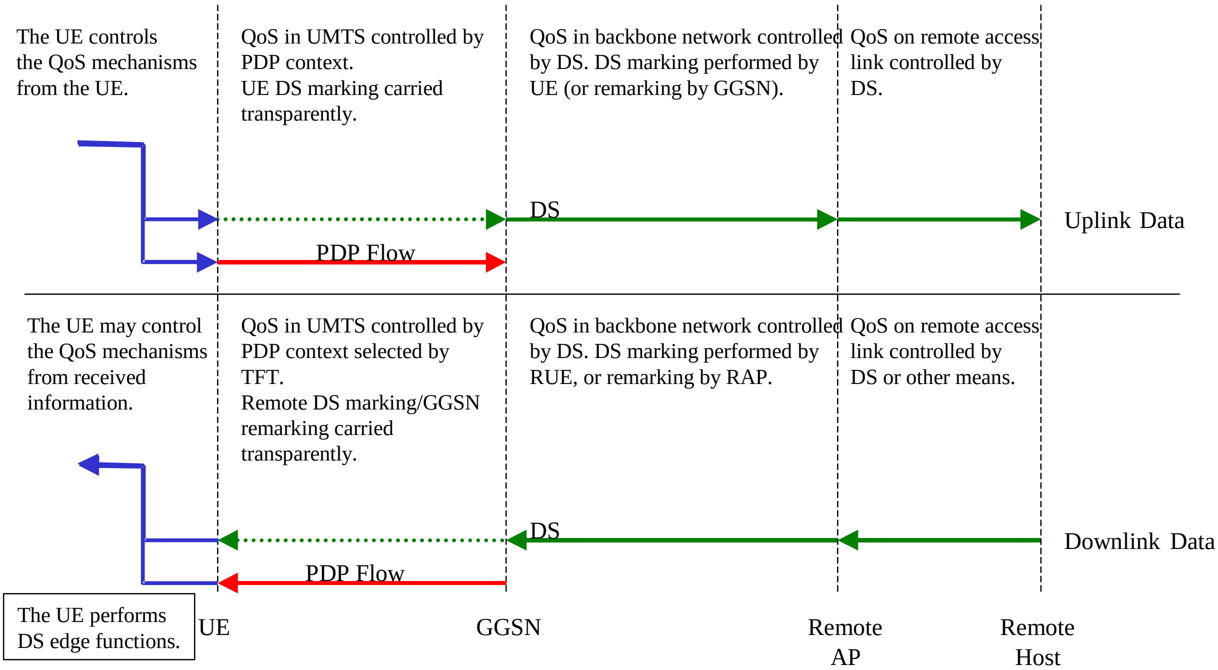

Figure A.3: Local UE supports DiffServ

A.2.3 Scenario 3

The UE performs an IP BS function which enables end-to-end QoS using IP layer signalling towards the remote end. There is no IP layer signalling between the IP BS Managers in the UE and the GGSN. However, the GGSN may make use of information regarding the PDP context which is signalled between the UMTS BS managers and provided through the translation/mapping function.

This scenario assumes that the UE and GGSN support DiffServ edge functions, and that the backbone IP network is DiffServ enabled. In addition, the UE supports RSVP signalling which interworks within the UE to control the DiffServ.

In this scenario, the control of the QoS over the UMTS access network (from the UE to the GGSN) may be performed either from the terminal using the PDP context signalling. Alternatively, subscription data accessed by the SGSN may override the QoS requested via signalling from the UE.

In this scenario, the terminal supports signalling via the RSVP protocol to control the QoS at the local and remote accesses, and DiffServ to control the IP QoS through the backbone IP network. The RSVP signalling protocol may be used for different services. It is only expected that RSVP using the Integrated Services (IntServ) semantics will be supported. It is only expected that only RSVP using the Integrated Services (IntServ) semantics would be supported, although in the future, new service definitions and semantics may be introduced. The entities that are supporting the RSVP signalling may fully support the specifications for IntServ and IntServ/DiffServ interwork. If not, they are expected to set the break bit.

The QoS for the wireless access is provided by the PDP context. The UE may control the wireless QoS through signalling for the PDP context. The characteristics for the PDP context may be derived from the RSVP signalling information, or may use other information.

QoS for the IP layer is performed at two levels. The end-to-end QoS is controlled by the RSVP signalling. Although RSVP signalling can be used end-to-end in the QoS model, it is not necessarily supported by all intermediate nodes. Instead, DiffServ is used to provide the QoS throughout the backbone IP network.

At the UE, the data is also classified for DiffServ. Intermediate QoS domains may apply QoS according to either the RSVP signalling information or DiffServ mechanisms. In this scenario, the UE is providing interworking between the RSVP and DiffServ domains. The GGSN may override the DiffServ setting from the UE. This GGSN may use information regarding the PDP context in order to select the appropriate DiffServ setting to apply, as shown in the figure below.

The end-to-end QoS is provided by a local mechanism in the UE, the PDP context over the UMTS access network, DiffServ through the backbone IP network, and DiffServ in the remote access network in the scenario shown in Figure A.4 below. The RSVP signalling may control the QoS at both the local and remote accesses. This function may be used to determine the characteristics for the PDP context, so the UE may perform the interwork between the RSVP signalling and PDP context.

The UE provides control of the DiffServ (although this may be overwritten by the GGSN), and in effect, determines the appropriate interworking between the PDP context and DiffServ.

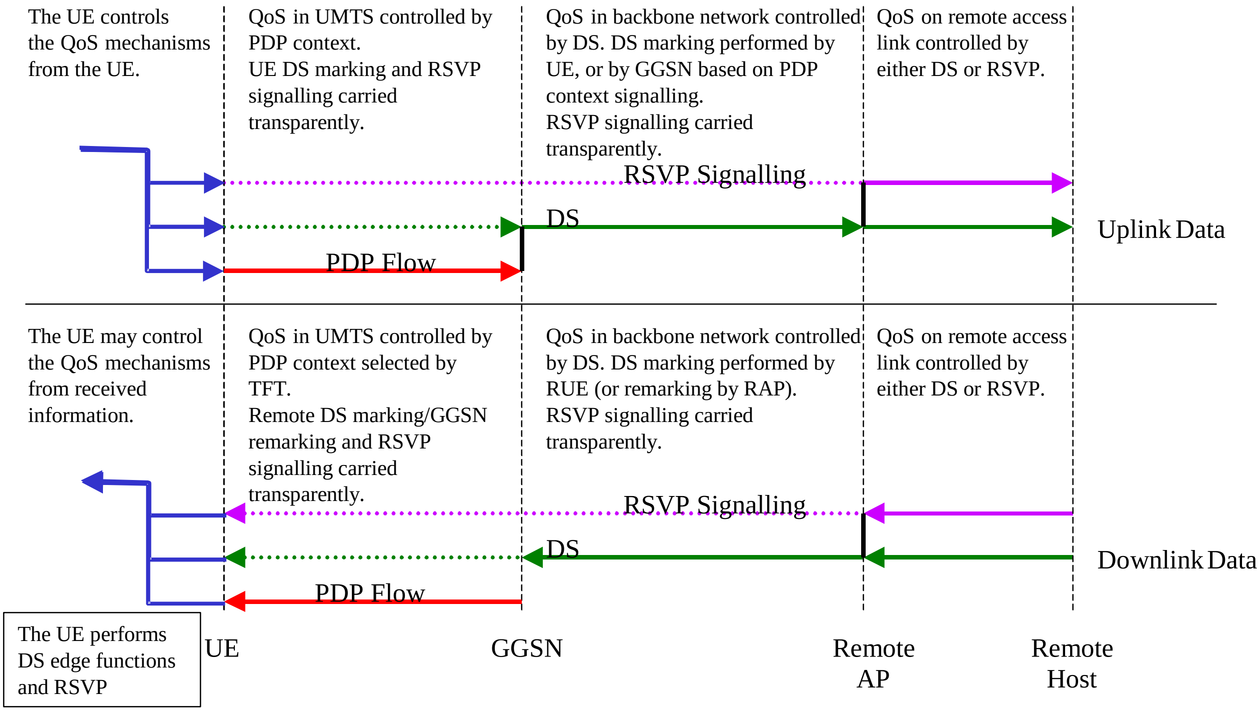

Figure A.4: Local UE supports RSVP signalling with IntServ semantics, and DiffServ

A.2.4 Scenario 4

The UE performs an IP BS function which enables end-to-end QoS using IP layer signalling towards the remote end. However, the UE relies on this end-to-end communication being utilised by at least the access point (GGSN) in order to provide the end-to-end QoS.

This scenario assumes that the UE and GGSN support RSVP signalling which may control the QoS directly, or interwork with DiffServ. The backbone IP network is RSVP and/or DiffServ enabled.

In this scenario, the terminal supports signalling via the RSVP protocol to control the QoS across the end-to-end path. The GGSN also supports the RSVP signalling, and uses this information rather than the PDP context to control the QoS through the backbone IP network. The control of the QoS through the core is expected to be supported through interworking with DiffServ at the GGSN, although it may optionally be supported by per flow resource reservation. The RSVP signalling protocol may be used for different services. It is only expected that only RSVP using the Integrated Services (IntServ) semantics would be supported, although in the future, new service definitions and semantics may be introduced. The entities that are supporting the RSVP signalling may fully support the specifications for IntServ and IntServ/DiffServ interwork. If not, they are expected to set the break bit.

In this scenario, the control of the QoS over the UMTS access network (from the UE to the GGSN) may be performed either from the terminal using the PDP context signalling. Alternatively, subscription data accessed by the SGSN may override the QoS requested via signalling from the UE.

QoS for the IP layer is performed at two levels. The end-to-end QoS is controlled by the RSVP signalling. Although RSVP signalling occurs end-to-end in the QoS model, it is not necessarily supported by all intermediate nodes. DiffServ is used to provide the QoS throughout the backbone IP network, although optionally each node may support RSVP signalling and allocation of resources per flow.

The GGSN supports the RSVP signalling and acts as the interworking point between RSVP and DiffServ. Intermediate QoS domains may apply QoS according to either the RSVP or DiffServ mechanisms.

The end-to-end QoS is provided by a local mechanism in the UE, the PDP context over the UMTS access network, DiffServ through the backbone IP network, and RSVP in the remote access network in the scenario shown in Figure A.5 below. The RSVP signalling may control the QoS at the local access. This function may be used to determine the characteristics for the PDP context, so the UE may perform the interwork between RSVP and the PDP context.

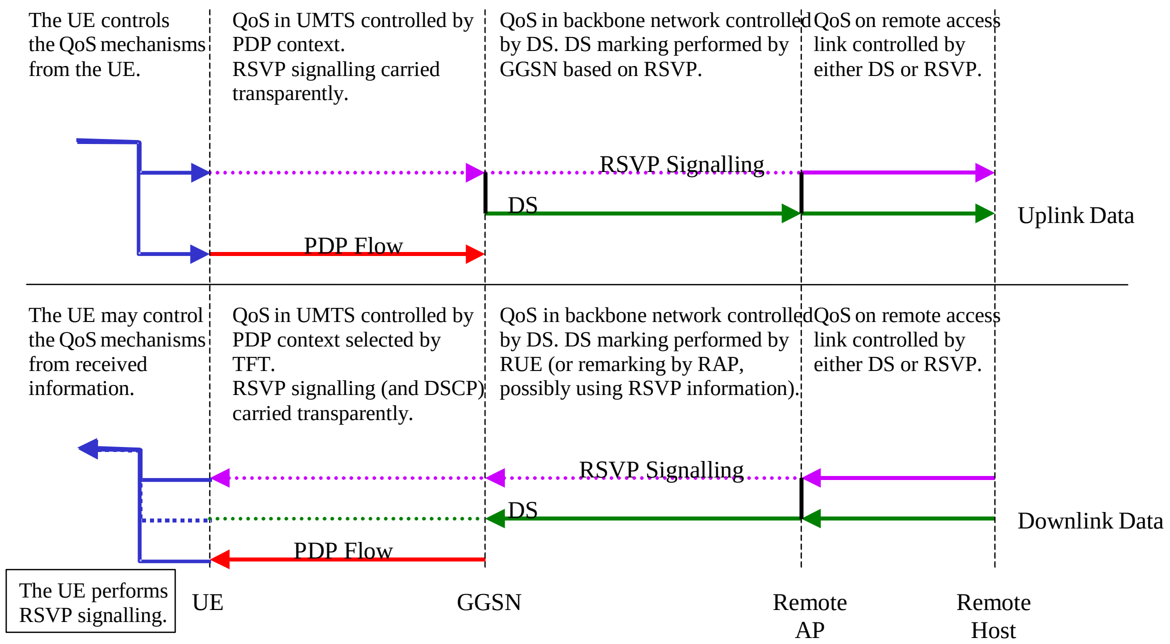

Figure A.5: Local UE supports RSVP signalling using IntServ Semantics

A.2.5 Scenario 5

The UE performs an IP BS function which enables end-to-end QoS without IP layer signalling and negotiation towards the IP BS function in the GGSN, or the remote host. The UE provides IP specific information to the GGSN, using IP specific elements of the PDP context activation/modification message, to enhance the interworking options to the DiffServ edge function of the GGSN.

The scenario assumes that the GGSN support DiffServ edge functions, and that the backbone IP network is DiffServ enabled.

The GGSN DiffServ edge function may use the IP specific information for the DiffServ classifier functionality, e.g., 5-tuple combination of source and destination IP address, source and destination port number, and the protocol identifier. The information can also be used for DiffServ class admission control, e.g., the requested end-to-end bandwidth from the UE for a particular flow may be informed to the GGSN beforehand for the GGSN DiffServ edge to determine if the flow can be allowed to a certain DiffServ class or an egress point. As a result, the GGSN may select the appropriate DiffServ setting to apply. This is shown in the figure below.

There exist IP specific elements of PDP context activation/modification message that are transferred from the UE to the GGSN. The elements are discussed in Appendix D.

In this scenario, the control of the QoS over the UMTS access network (from the UE to the GGSN) may be performed from the terminal using the PDP context signalling. Alternatively, subscription data accessed by the SGSN may override the QoS requested via signalling from the UE.

TheQoS for the downlink direction is controlled by the remote host from the remote network to the GGSN. The PDP context controls the UMTS level QoS between the GGSN and the UE. The QoS in the uplink direction is controlled by the PDP context up to the GGSN. The GGSN uses the IP specific elements of UMTS signalling to interwork with DiffServ in the backbone IP network and controlling the IP QoS bearer service towards the remote -host.

The end-to-end QoS is provided by a local mechanism in the UE, the PDP context over the UMTS access network, DiffServ through the backbone IP network, and DiffServ in the remote access network. Note that DiffServ control at the Remote Host is shown in this example. However, other mechanisms may be used at the remote end, as demonstrated in the other scenarios.

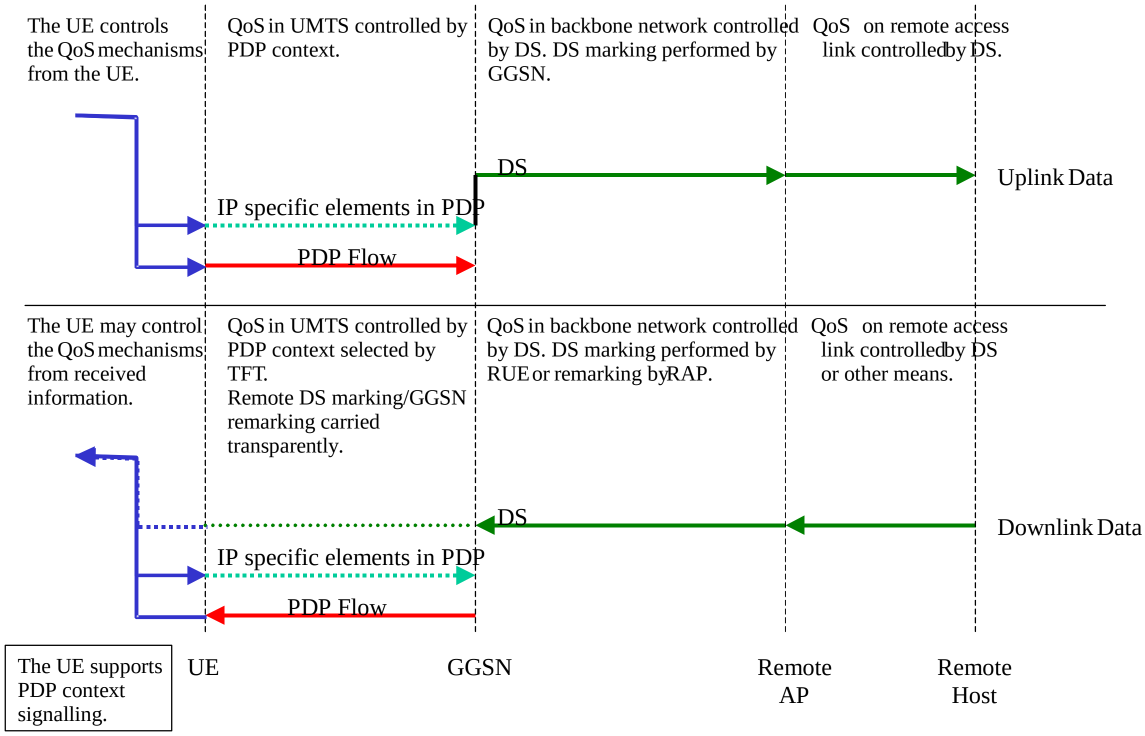

Figure A.6: Local UE supports IP specific elements in PDP context activation/modification message and GGSN provides interworking with DiffServ

A.2.6 Scenario 6

The UE performs an IP BS function which enables end-to-end QoS without IP layer signalling and negotiation towards the IP BS function in the GGSN, or the remote host. The UE provides IP level end-to-end QoS information to the GGSN, using IP specific elements of the context activation/modification message , to enhance the interworking options to an RSVP function in the GGSN. The end-to-end IP QoS bearer service towards the remote host is controlled from the GGSN.

The scenario assumes that the GGSN supports DiffServ edge functions, and the backbone IP network is DiffServ enabled. This scenario does not preclude the backbone IP network from having RSVP non-transparent routers.

The GGSN may use the IP specific elements in PDP context activation/modification messageto invoke RSVP messages to setup the uplink as well as the downlink flows in the backbone IP network up to the remote host. For example, in the uplink direction, the GGSN may use the IP specific elements in PDP context activation/modification message to generate the RSVP Path messages, with the desired QoS / traffic specification, to the specified destination IP address. Also, the GGSN DiffServ edge function may use the IP specific elements in PDP context activation/modification messageto select the appropriate DiffServ setting to apply. This is shown in the figure below.

There exist IP specific elements of PDP context activation/modification message that is transferred from the UE to the GGSN. The elements are discussed in Appendix D.

In this scenario, the control of the QoS over the UMTS access network (from the UE to the GGSN) may be performed either from the terminal using the PDP context signalling. Alternatively, subscription data accessed by the SGSN may override the QoS requested via signalling from the UE.

The QoS for the downlink direction is controlled by the PDP context between the UE and the GGSN. The GGSN terminates the RSVP signalling received from the remote host, and may use the information in theIP specific elements in PDP context activation/modification message when processing RSVP. The QoS in the uplink direction is controlled by the PDP context up to the GGSN. The GGSN may use the IP specific elements in PDP context activation/modification message to provide the interworking with RSVP towards the remote host. The IP specific elements in PDP context activation/modification message may allow for the establishment of the RSVP session..

The end-to-end QoS is provided by a local mechanism in the UE, the PDP context over the UMTS access network, DiffServ through the backbone IP network, and RSVP in the remote access network.

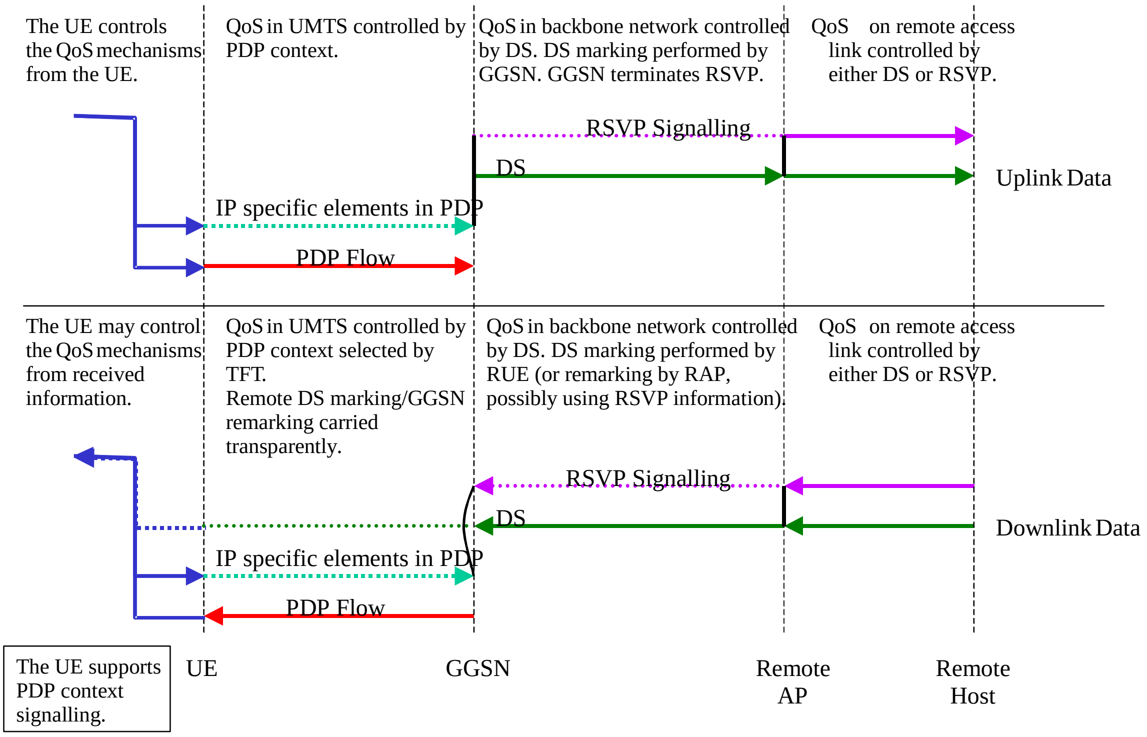

Figure A.7: Local UE supports IP specific elements in PDP context activation/modification message and GGSN provides interworking with RSVP

A.3 RSVP Usage for End-to-End QoS in UMTS

This section contains typical RSVP usage for end-to-end QoS which is applicable to UMTS networks. It aims to convey the general cases when RSVP is used as IP layer signalling protocol from the UMTS network to the remote end terminal to enable end-to-end QoS.

Clarification of terminology:

RSVP transparent – describes the case where GGSN is transparently relaying the RSVP messages, i.e., the GGSN does not process any RSVP message.

RSVP non transparent – describes the case where the GGSN processes RSVP message. (For example the GGSN may maintain RSVP soft states, and/or may use QoS information derived to interwork with other QoS mechanisms (e.g., DiffServ, PDP context), and may forward RSVP messages onwards.)

NOTE: It is assumed that there is an existing PDP context that carries signalling (e.g., SIP or RSVP) between the UE and GGSN.

A.3.1 RSVP in Scenarios 3 and 4

In Scenarios 3 and 4, the UE is supporting RSVP enabled applications. In order to provide QoS over the UMTS segment in an appropriate manner for the end-to-end IP QoS requested through RSVP, the UE shall examine the RSVP signalling and use this information to control the activation/modification of the PDP context as shown in Figures A.8 and A.9 below. The UE must not reply with RSVP RESV before the PDP context resources is setup.

Figures A.8 and A.9 do not preclude the case where the GGSN is RSVP transparent as shown in Scenario 3.

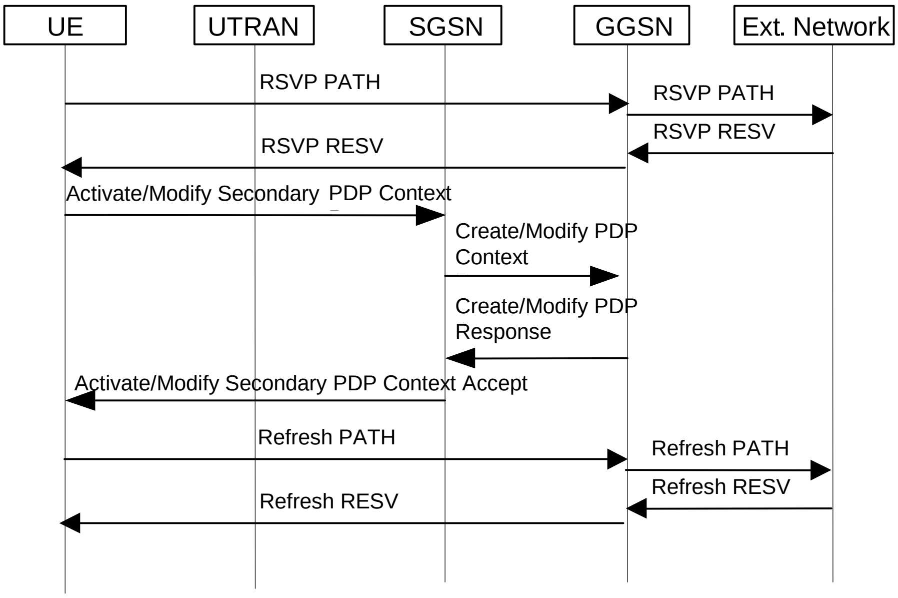

Figure A.8: UE supports RSVP signalling for the uplink flow

NOTE The above diagram depicts only one possible signalling sequence, depending on UE implementation decision, other alternative signalling sequences are:

- to trigger the Activate/Modify Secondary PDP Context when the PATH message is generated by the UE (either directly or after some timeout) without waiting for the arrival of the RESV message.
- to trigger the Activate/Modify Secondary PDP Context before the PATH message is forwarded by the UE.

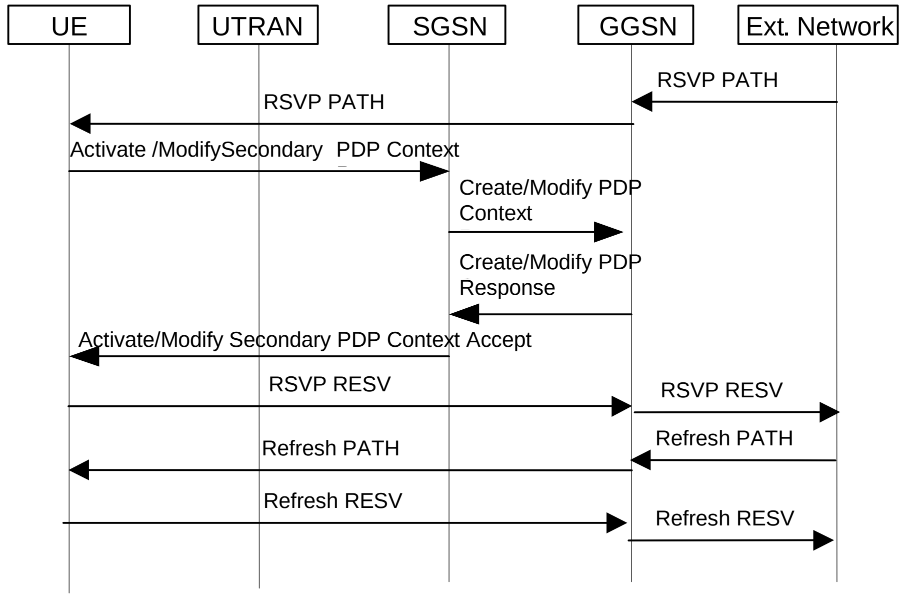

Figure A.9: UE supports RSVP signalling for the downlink flow

A.3.2 RSVP in Scenario 6

In Scenario 6, the UE provides IP level end-to-end QoS information to the GGSN using IP specific elements in PDP activation/modification message, and the GGSN uses this information to invoke RSVP messages to setup the uplink as well as the downlink flows. RSVP signalling is generated and terminated by the GGSN as shown in Figures A.10 and A.11 below.

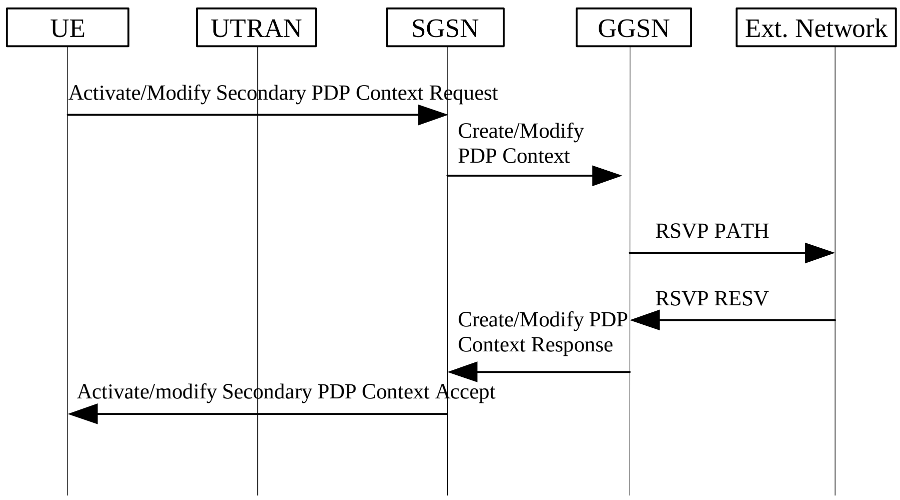

Figure A.10: UE provides IP specific elements in PDP activation/modification message, and GGSN invokes RSVP messages for uplink flow

[Editorial Note: the above diagram depicts a signalling sequence, however, the alternative signalling sequences below are possible and are for further study:

- *to trigger the Create PDP Context Response message on the PATH or after a timeout without waiting for the RESV message.*

- *to trigger the Create PDP Context Response message before the PATH.]*

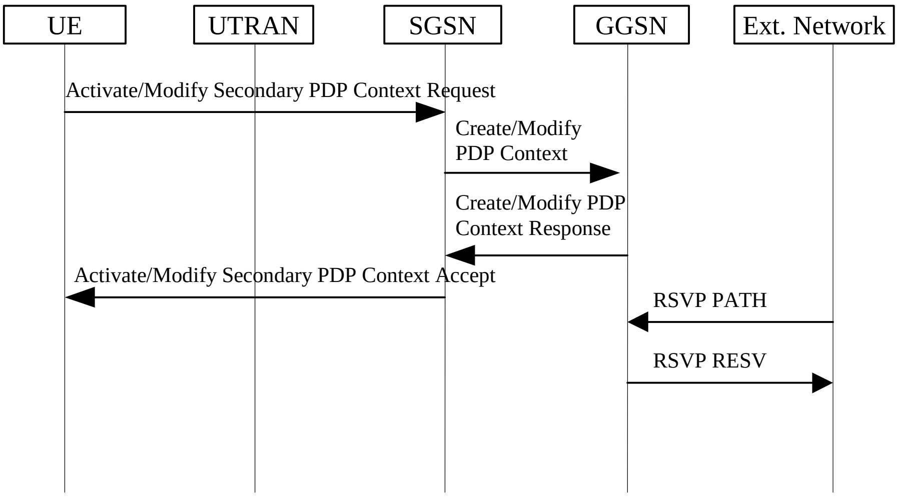

Figure A.11: UE provides IP specific elements in PDP activation/modification message, and GGSN terminates and invokes RSVP messages for downlink flow

In the above figure, the association between the RSVP PATH message and the PDP context is carried out in the GGSN.

[Editorial Note: the above diagram depicts a signalling sequence, however, the alternative signalling sequence below is possible and is for further study:

- *the RSVP PATH message arrives in the GGSN before the Activate/Modify Secondary PDP Context Request from the UE is received in the GGSN.]*

Annex B (Informative): IP Specific Elements in PDP Context Activation and Modification Message

[Editorial Note: The details of the IP specific elements in PDP context activation/modification message needs further study within the S2 QoS drafting group.]

If an IP BS Manager exists both in the UE and the GGSN, it is possible that these IP BS Managers communicate directly with each other by using relevant signalling protocols, e.g., RSVP. However, it is forseen that low end mobiles will not be able to support RSVP signalling, nevertheless these mobiles are required to be able to support end-to-end QoS requests from the application layer.

It is conceivable that there is a need to allow IP specific elements to be passed from the UE to the GGSN, between peer IP BS managers, without the necessity for signalling protocol support in the UE.

Figure B.1 below shows the two QoS control levels (the UMTS bearer level and IP bearer level) and is an example of how the IP BS manager at the GGSN may exercise admission control. The requested IP level QoS bandwidth received in the IP specific element from the UE for a particular flow may be informed to the GGSN beforehand for the GGSN DiffServ edge to determine if the flow can be allowed to a certain DiffServ class or an egress point based on the service level agreement (SLA).

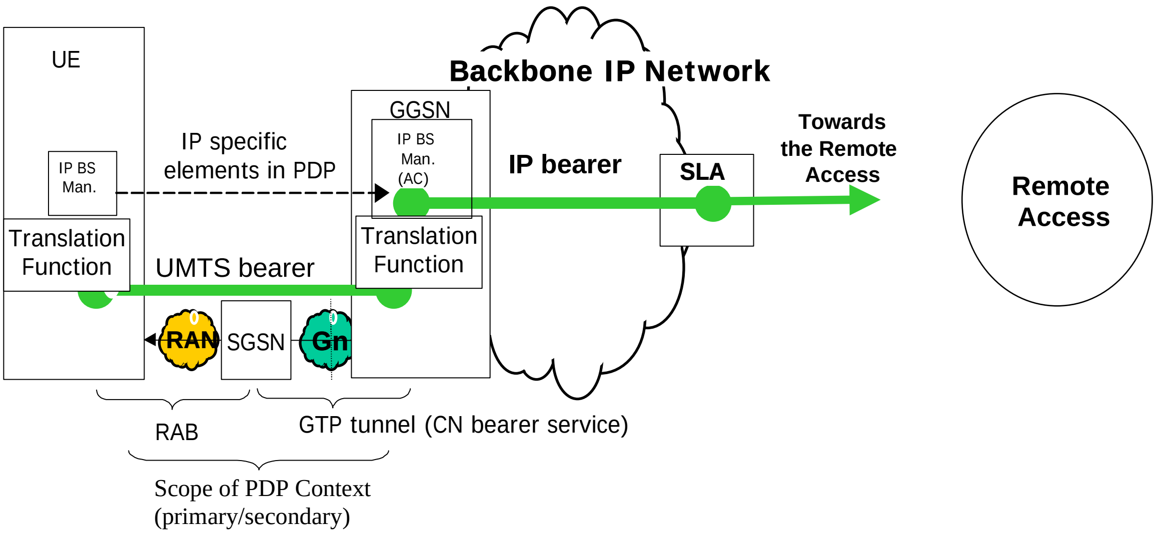

Figure B.1: Providing end-to-end QoS by means of transferring IP specific elements in PDP context activation/modification message between the UE and the GGSN and exercising policy enforcement / admission control in the GGSN

Within the IP policy architecture applicable to UMTS, the IP Policy Control makes decisions in regard to network based IP policy using policy rules, and communicates these decisions to the IP BS Manager in the GGSN, which is the IP policy enforcement point. Enforcement of policy may cover, among other things, the following requirements:

1. Authorisation of UMTS bearers from the application.
2. Control of opening and closing the gate for data to enter the network, controlled from the application server through the policy server.
3. Control of the level and destination of data permitted to pass the gate and enter the network, controlled from the application server through the policy server.

In order for the GGSN to enforce policy conformant to the requirements above there is a need to transfer information which belongs to the IP level between the peer IP BS manager entities in the UE and GGSN. To facilitate clean separation between the UMTS bearer level and the IP bearer level, this transfer of IP level information should be carried out in a manner that is transparent to the UMTS BS managers. The IP level information may include, for example in the RSVP case, the traffic flow specification contents, required IP level QoS, and destination IP address as described in

Scenario 6. For the DiffServ case, QoS information for the DiffServ classifier functionality and DiffServ class admission control may be included as described in Scenario 5.

The following candidates for the IP specific elements in PDP context activation/modification messageare identified:

1) Optional IP specific information is carried in the PDP context transparent to the UMTS BS managers, between peer IP BS managers, from the UE to the GGSN. The PDP context may contain a set of QoS attributes: "UMTS Specific IP QoS Attributes."

2) Extending the Traffic Flow Template concept to convey information to the GGSN related to the uplink flow (new mechanism), together with the use of existing TFT mechanism for the downlink flow. Bandwidth requirements etc. may be obtained by mapping the UMTS QoS parameters to the IP level QoS parameters, and also possibly extending the UMTS QoS parameters.

Annex C (Informative): IP BS Manager Functionality in the GGSN

The following will be taken as a working assumption for further work.

The IETF Differentiated Services architecture will be used to provide QoS for the external bearer service. The GGSN provides the Diffserv edge function as is currently described in 23.207.

For applications that require stringent IP quality of service (e.g., carrier-grade telephony),

1. It is intended to make possible the use of existing QoS management mechanisms in 23.107 with any required extensions FFS (e.g., IP bearer service information element) to request end-to-end QoS, and

2. When these existing QoS management mechanisms are used for requesting end-to-end QoS, the IP BS manager in the GGSN will act as an RSVP endpoint towards the external network, generating and responding to per-flow RSVP messages.

Use of RSVP is intended to enable the external network provider to support traffic engineering, efficient resource management, and call blocking if needed to handle temporary overload conditions.

It is also desired to allow endpoints to use Diffserv or RSVP to indicate their QoS requirements to the MT. The requirements on the IP bearer service manager in the UE and GGSN to support this functionality are FFS.

At PDP context setup the user shall have access to one of the following alternatives :

- Basic GPRS IP connectivity service: The bearer is established according to the user’s subscription, local operator’s IP bearer resource based policy, local operator´s admission control function and GPRS roaming agreements. In this case, IP bearer resource based local policy decisions may be applied to the bearer.

- Enhanced GPRS based services : The bearer is used to support an enhanced application-layer service, such as IM. In this case, service-based local policy decisions (e.g., authorization and gating of the bearer by a proxy CSCF)) may be applied to the bearer.

C.1 Policy Enforcement Point in GGSN

The following is taken as the basis for further work.

This section provides a functional definition of the Policy Enforcement Point (PEP) implemented in the IP bearer service manager in the GGSN in order to meet UMTS architectural requirements.

The Policy Enforcement Point (PEP) controls access to quality of service for a set of IP packets that match a packet classifier. Policy decisions are either “pushed” to the GGSN by a policy control function, or alternatively, the GGSN may request policy information from a policy control function on receipt of an IP bearer resource request.

Policy enforcement is defined in this section in terms of a “gate” implemented in the GGSN. A gate is a policy enforcement function for a unidirectional flow of packets, e.g., in either the upstream or downstream direction. At a high level, a gate consists of a packet classifier, a resource “envelope,” and an action taken when the set of packets matching the classifier exceeds the resource envelope.

Unidirectional gates are used since the basic unit of IP bearer service resource allocation is for unidirectional flows. For example, a downstream-only gate would be used when an application on a UE has subscribed in a receive-only mode to an IP multicast session. A downstream-only flow is also needed to support "remote ringback" in an IP telephony application, where the ringback signal is generated remotely by a PSTN gateway or remote UE. For this application, to avoid certain theft of service scenarios it is necessary to enable the downstream flow of packets to the session originator, while not enabling the upstream flow of packets until the remote UE picks up. When access to QoS for a flow of packets in both directions is desired, a pair of gates is used.

The information that may be associated with a gate is described below. This information is not necessarily the only information that might be used, but is intended to cover the currently understood applications. A gate is described by the following information:

- Packet classifier
- Authorized envelope
- Action
- Resource identifier
- Reserved envelope
- Gate identifier

The *packet classifier* associated with a gate is described by the following information:

- Direction
- Source IP address
- Destination IP address
- Source port
- Destination port
- Protocol

The direction indicates whether the gate is an upstream or downstream gate. The source IP address and port identify the source IPv4 or IPv6 address and port, as seen at the GGSN. The destination IP address and port identify the destination IPv4 or IPv6 address and port, as seen at the GGSN. The protocol field identifies the IP protocol type of the packet. With the exception of the direction, these fields can be wild-carded. For example, in a SIP session, the source port for the bearer is not exchanged in SIP signaling messages, and therefore cannot be set up when the gate is initialized.

The *authorized envelope* defines an upper bound or "envelope" of the resources that are authorized for the set of packets defined by the packet classifier. The authorized envelope can authorize more resources than are actually used. Since the authorized envelope defines IP bearer resources towards or from the external network, it is appropriate to express it in terms of IP bearer resources such as a peak information rate, mean information rate, and token bucket size to or from the external network. For example, an Intserv Flowspec is an appropriate representation of IP bearer resources. The authorized envelope is mapped to (a range of) UMTS bearer resources by the translation function in the GGSN when necessary. The authorized envelope allows the PCF to pre-authorize a flow, before the UE requests allocation of the resources (“push” model), as shown in TS23.228.

The *action* defines the action to be taken when the set of packets defined by the packet classifier exceeds the authorized envelope (or reserved envelope, below). The action includes marking out-of-profile packets with a particular Diffserv Code Point (DSCP), marking in-profile packets with a particular DSCP, shaping to a token bucket, or packet dropping.

According to the above definitions, a set of packets may match more than one classifier. When this happens, the actions associated with the each of the applicable gates are considered to be executed in sequence, in the order in which the gates were configured in the GGSN. Packets that are marked by a gate may not be (re)marked by a subsequent gate to a DSCP corresponding to a better service class.

The *resource identifier* identifies a set of resources that can be shared by multiple gates, e.g., for several sessions. For example, the resource identifier might allow a UE to share a single set of resources for two sessions that do not simultaneously use the resources, such as during call waiting. The resource identifier would be included in the IP bearer service information element of a PDP context activation/modification request to support this function.

The *reserved envelope* defines an upper bound or "envelope" of the resources that are reserved for a bearer. The reserved envelope is advantageous if it is possible for a UE to reserve more resources than are actually used. For example, during call waiting, the UE might maintain a reservation for a high quality codec for a temporarily inactive session, while using a lower quality codec for the active session. The reserved envelope implies that resource reservation performs admission control, and if successful, sets aside the requested resources in a pre-emptible mode. Support for this functionality requires the PDP context activation and modification procedures to distinguish between reservation requests and resource use.

The *gate identifier* (GateID) uniquely identifies a gate at a GGSN. The GateID can be used to correlate resource reservation requests from the UE (e.g., PDP context activation request) with authorization commands from the PCF. To support this function, the GateID needs to be included in the PDP context activation request, the policy control interface, and SIP signaling messages between the proxy CSCF and UE. Details of this are described in Internet draft [<draft-ietf-sip-call-auth-00.txt>] "SIP Extensions for Media Authorization". It is convenient to allow a GateID to be associated with one or more gates: a single upstream gate, a single downstream gate, or an upstream and a downstream gate.

C.2 Policy Control Interface

The following is taken as the basis for further work.

IP QoS admission control manages allocation of QoS resources based on administrative policy and available resources. Admission control based on administrative policy is performed using a client/server architecture involving policy enforcement in the IP bearer service manager in the GGSN, and a policy control function (PCF).

When the Common Open Policy Service (COPS) protocol [RFC2748] is used as the client/server protocol between the PEP and the PCF, the COPS client (PEP) can request policy information from the PCF triggered by a QoS signaling request. Alternatively, policy decisions made by the PCF can be pushed to the COPS client (PEP) based on an external out-of-band QoS service request, e.g., triggered by SIP signaling. These policy decisions are stored in the COPS client in a local policy decision point accessed by the PEP to make admission control decisions without requiring additional interaction with the PCF.

The COPS protocol supports several messages between a client and server. These messages consist of the following operations that may be performed:

- Client-Open/Client-Accept/Client-Close
- Request
- Decision
- Report State
- Delete Request State
- Keep Alive
- Synchronize State Request/Synchronize State Complete.

This section addresses the policy interface to allow the PCF to push policy information to the PEP, consistent with the information flows defined in TS23.228. Information flows in which the PEP requests policy information from the PCF has not yet been considered. Support for this functionality in the policy interface is therefore left FFS.

Since we consider a push model, messages from PCF to the GGSN are COPS Decision messages, and messages from GGSN to PCF are COPS Report messages.

C.2.1 Additional COPS Objects Needed for Policy Control

Additional information elements need to be included in COPS messages to support the UMTS QoS architecture. Consistent with the COPS framework, the policy control interface is identified by a unique "client type" allocated for a UMTS client (GGSN). The remainder of this section describes the objects that are currently identified as being needed in the architecture:

- gate identifier
- command/response
- gate spec
- event generation info
- endpoint identifier
- max gates

The *gate identifier* contains a unique identifier for the gate that is being referenced in the command or response. To support multiple policy control functions that may be providing policy control of a GGSN, the gate identifier should be allocated by the GGSN.

The *commands* that are used from the PCF to the PEP include commands to:

- allocate a gate (Gate-Alloc)
- set parameters associated with a gate (Gate-Set)
- delete a gate (Gate-Delete)
- get parameters/information associated with the gate (Gate-Info)

The *responses* that are needed from the PEP to the PCF include an acknowledgement and/or an error response to each of these commands, e.g., Gate-Alloc-Ack, Gate-Alloc-Err, etc.

The *gate spec* object contains the specification of the gate parameters that are being set or returned in a response.

- Direction
- Source IP address
- Destination IP address
- Source Port
- Destination Port
- Protocol
- Action
- DSCP Field
- Flowspec (or Flowspecs)

The *event generation info* contains information related to usage recording that may be needed for IP QoS bearers. This might include a "billing identifier" needed to correlate event records from the GGSN with event records from the proxy CSCF, so that all records associated with the same session can be associated. In order for the billing identifier to be unique, it might include for example, a long numeric value generated by the PCF, along with the identity of the PCF.

The *endpoint identifier* and *max gates* fields are used to prevent a UE-initiated denial of service attack that attempt to set up an excessive number of simultaneous sessions, resulting in the allocation of multiple gates. The endpoint identifier contains the identity (e.g., IP address) of the endpoint associated with the gate, while the max gates field contains the maximum number of gates that can be allocated to this particular endpoint. The GGSN can return an error if the number of allocated gates exceeds max gates.

Annex D (Informative): Mechanisms for Interaction with External Networks

If the UMTS network must interact with an external network for IP resource management, this may be performed in a number of ways, including:

[Editorial Note: This list is not exhaustive. Other options have to be investigated.]

- Signalling along the flow path. In this scenario, resource requirements are explicitly requested and either granted or rejected through the exchange of signalling messages between network elements along the path of the IP packet flow. Signalling may be performed on a per-flow basis (e.g. using end to end RSVP) or it may be performed for an aggregate set of flows; it the latter case, it is expected that signalling exchanges would only be required when there are significant changes required in the resources allocated to an aggregate set of flows.

- Interaction between network management entities. In this scenario, resource requirements are explicitly requested and either granted or rejected through the exchange of signalling messages between network management entities. The results of this exchange are then communicated to the affected network elements along the path of the flow through policy or network management transactions (e.g. COPS, SNMP).

- Service Level Agreements enforced by the border routers between networks. In this scenario, resources are allocated along the path based on à priori agreements between the network operators. The border routers along the path flow are provisioned with the characteristics of the aggregated traffic that is allowed to flow between systems.

Annex E (informative): Change History

Change history				
Date	**TSG #**	**TSG Doc.**	**Rev**	**Subject/Comment**
2000-10		23.XXX	0.0.0	Initial version of the specification presented at S2 QoS Drafting Meeting in Sophia Antipolis
2000-10		23.207	0.0.1	Revision with comments from the S2 QoS Drafting Meeting in Sophia Antipolis. Updates to TS title, scope. TS number included.
2000-11	S2#15	23.207	0.0.2	This version is not approved.
2000-11	S2#15	23.207	0.0.3	Annex C (IP BS Manager Functionality in the GGSN) is added. S2 QoS Group-approved editorial updates are incorporated which have been raised before and during the QoS Drafting Meeting at Makuhari, Japan.
2000-11	S2#15	23.207	0.1.0	This version incorporates all agreements that were made during QoS Drafting Meeting/S2#15 at Makuhari, Japan. Major changes are the following: Subsections 5.1.1.2 (Interactions to External Networks), C.1 (Policy Enforcement Point in GGSN), C.2 (Policy Control Interface), C.2.1 (Additional COPS Objects Needed for Policy Control), and Annex D (Mechanisms for Interaction with External Networks) are added. Text related to working assumption on interaction of bearer service with policy control is added to Annex C. Text related to application on the external device is added to Subsection 4.1
2000-11	S2#15	23.207	1.0.0	Content is the same as V0.1.0 (change marks have been cleaned). This version is submitted to SA#10 for information.

Editor: Johnson Oyama (Johnson.Oyama@era.ericsson.se)