

QoS Control in the 3GPP Evolved Packet System

Hannes Ekström, Ericsson

ABSTRACT

In this article we describe the QoS concept of the evolved packet system, which was standardized in 3GPP Release 8. The concept provides access network operators and service operators with a set of tools to enable service and subscriber differentiation. Such tools are becoming increasingly important as operators are moving from a single to a multi-service offering at the same time as both the number of mobile broadband subscribers and the traffic volume per subscriber is rapidly increasing.

The “bearer” is a central element of the EPS QoS concept and is the level of granularity for bearer-level QoS control. The network-initiated QoS control paradigm specified in EPS is a set of signaling procedures for managing bearers and controlling their QoS assigned by the network. The EPS QoS concept is class-based, where each bearer is assigned one and only one QoS class identifier by the network. The QCI is a scalar that is used within the access network as a reference to node-specific parameters that control packet forwarding treatment. This class-based approach, together with the network-initiated QoS control paradigm, gives network operators full control over the QoS provided for its offered services for each of its subscriber groups.

INTRODUCTION AND MOTIVATION

In recent years, cellular operators across the world have seen a rapid growth of mobile broadband subscribers. At the same time, the traffic volume per subscriber is also increasing rapidly; in particular, with the introduction of flat-rate tariffs and more advanced mobile devices. Operators are moving from a single-service offering in the packet-switched domain (Internet access) to a multi-service offering by adding new services that are also provided across the mobile broadband access. Examples of such services are multimedia telephony and mobile-TV. These services have different performance requirements, for example, in terms of required bit rates and packet delays. Solving these performance issues through over-provisioning typically is uneconomical due to the relatively high cost for transmission capacity in cellular access networks (including radio spectrum and backhaul from the base stations).

¹ The source and destination IP address, source and destination port number, and protocol ID typically are referred to as the IP five-tuple.

In addition, operators have started to provide subscriber differentiation, that is, differentiating the treatment received by different subscriber groups for the same service. These subscriber groups can be defined in any way that is suitable to the operator, for example, corporate versus private subscribers, post- versus pre-paid subscribers, and incoming roaming subscribers. Hence, there is a need to standardize simple and effective QoS mechanisms for multi-vendor mobile broadband deployments. Such QoS mechanisms should allow the access operator to enable service and subscriber differentiation and to control the performance experienced by the packet traffic of a certain service and subscriber group as depicted in Fig. 1.

This article presents the network-initiated and class-based concept for QoS control that was standardized for the evolved packet system (EPS). The basis and motivation for this concept was outlined in [1]. This article further describes the QoS mechanisms that are enabled in the EPS by the Third Generation Partnership Project (3GPP) Release 8 specifications.

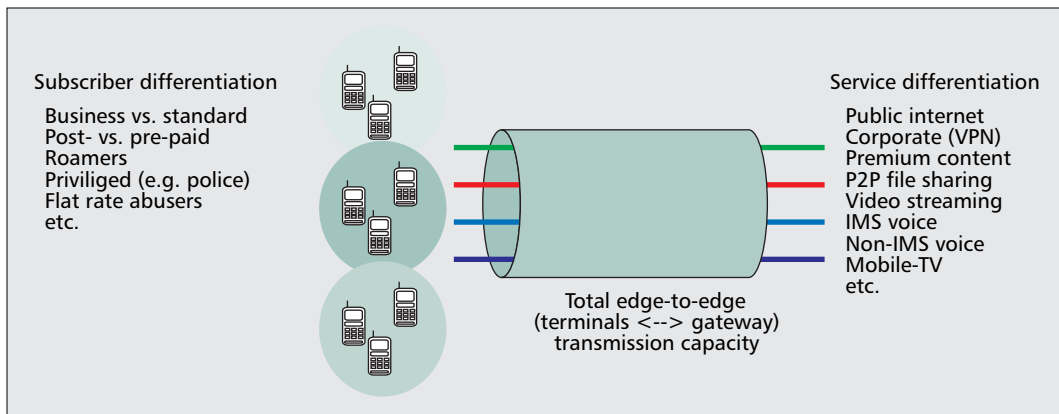
The article is organized as follows. In the next section, we describe the components of the EPS QoS concept, and we then describe the QoS paradigms standardized for EPS, the network-initiated and terminal-initiated, and further describe the benefits of using the network-initiated paradigm. The following section provides an example of an end-to-end use case of providing service and subscriber differentiation using the EPS QoS concept. The final section concludes the article.

THE EPS QoS CONCEPT

In this section, we describe the details of the EPS bearer, its associated QoS parameters, and the EPS QoS mechanisms that are enabled by the standard.

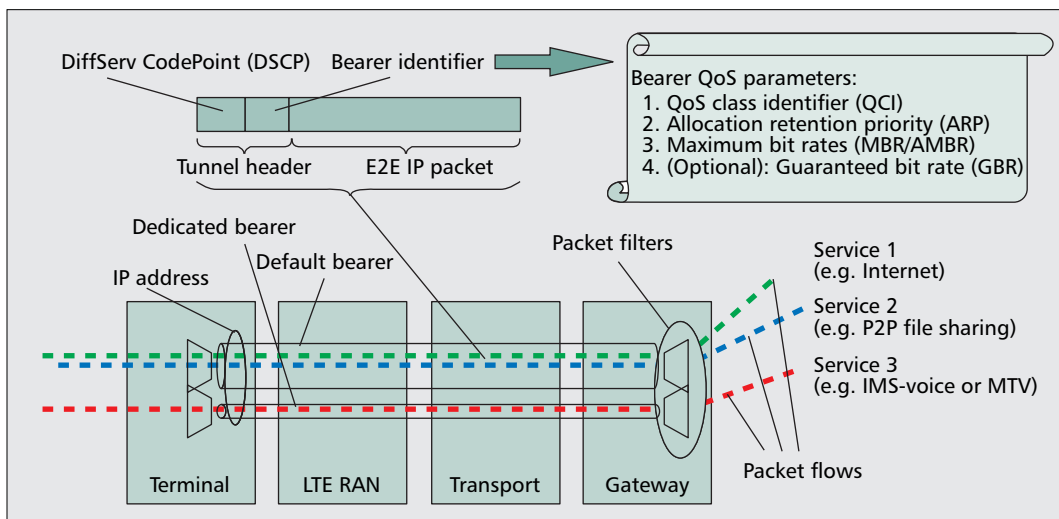
THE BEARER

An EPS bearer — “bearer” for short — uniquely identifies packet flows that receive a common QoS treatment between the terminal and the gateway. A packet flow is defined by a five-tuple-based¹ packet filter, that is, the packet filters in the terminal (for uplink traffic) and the gateway (for downlink traffic) determine the packet flows associated with an EPS bearer (Fig. 2).



■ **Figure 1.** Providing service and subscriber differentiation.

The bearer is the basic enabler for traffic separation, that is, it provides differential treatment for traffic with differing QoS requirements. The concept of the bearer and the associated signaling procedures further enable the system to reserve system resources before packet flows that are mapped to that bearer are mapped into the system.



■ **Figure 2.** The bearer and its associated QoS parameters.

A bearer is the level of granularity for bearer-level QoS control in the EPS. That is, all packet flows mapped to the same bearer receive the same packet-forwarding treatment (e.g., scheduling policy, queue management policy, rate-shaping policy, link-layer configuration, etc.). Providing different packet-forwarding treatment requires separate bearers.

One bearer exists per combination of QoS class and IP address of the terminal. The terminal can have multiple IP addresses, for example, in case it is connected to multiple access point names (APNs, one IP address per APN). The APN is a reference to the IP network to which the system connects the terminal. That is, the terminal can have two separate bearers associated with the same QoS class to two different APNs.

Each end-to-end IP packet entering the system is provided with a tunnel header on the different system interfaces. This tunnel header contains the bearer identifier so that the network nodes can associate the packet with the correct QoS parameters. In the transport network, the tunnel header further contains a diffserv code point (DSCP) value, as shown in Fig. 2.

The bearer is the basic enabler for traffic separation, that is, it provides differential treatment for traffic with differing QoS requirements. The

concept of the bearer and the associated signaling procedures (see later in this section and in the next section) further enable the system to reserve system resources (e.g., processing and transmission capacity) before packet flows that are mapped to that bearer are admitted into the system. The latter is performed through an admission control function that operates on a per-bearer level.

GBR vs. Non-GBR Bearers — Two types of bearers exist: guaranteed bit-rate (GBR) and non-guaranteed bit-rate (non-GBR) bearers. Provided that the traffic carried by a GBR bearer conforms to the value of the GBR QoS parameter associated with the bearer (discussed later in this section), the service(s) utilizing that GBR bearer can assume that congestion-related packet losses (i.e., packet losses caused by overflowing buffers) will not occur. This is realized by admission control functions that may reside in different network nodes (e.g., the long-term evolution [LTE] base station) and are executed at the point in time when a bearer becomes established or modified. A service utilizing a non-GBR bearer on the other hand, must be prepared to experience congestion-related packet loss.

A GBR bearer typically is established “on demand” because it blocks transmission resources by reserving them in an admission control function. On the other hand, a non-GBR bearer can remain established for long periods of time because it does not block transmission resources.

A GBR bearer typically is established “on demand” because it blocks transmission resources by reserving them in an admission control function. On the other hand, a non-GBR bearer can remain established for long periods of time because it does not block transmission resources.

An operator would choose GBR bearers for services where the preferred user experience is “service blocking over service dropping,” that is, block a service request rather than risk degraded performance of an already admitted service request. This is relevant in scenarios where it may not be possible to meet the demand for those services with the dimensioned capacity (e.g., in situations with extreme network load, like New Year’s Eve). Whether a service is realized based on GBR or non-GBR bearers is, therefore, an operator policy decision that to a large extent depends on expected traffic load versus dimensioned capacity. Assuming sufficiently dimensioned capacity, any service, both real time and non-real time, can be realized based on non-GBR bearers.

Default vs. Dedicated Bearers — Orthogonal to being classified as GBR or non-GBR, a bearer is either a default or a dedicated bearer. The default bearer is the bearer that is set up when the terminal attaches to the network. One default bearer exists per terminal IP address, and it is kept for as long as the terminal retains that IP address. The default bearer provides the basic connectivity. Because a default bearer can remain established for long periods, the 3GPP specifications mandate that the default bearer is a non-GBR bearer. The QoS level of the default bearer is assigned based on subscription data.

To provide different QoS in the network to two different packet flows for the same IP address of a terminal, one or more dedicated bearers are required. The dedicated bearer can be either a non-GBR or a GBR bearer. The operator can control which packet flows are mapped onto the dedicated bearer, as well as the QoS level of the dedicated bearer through policies that are provisioned into the network policy and charging resource function (PCRF) [2]. In this article, we refer to that node simply as the policy controller. Figure 2 shows a terminal with a default and a dedicated bearer established to the same terminal IP address.

The policy controller defines specific packet flows to be mapped onto a dedicated bearer and typically defines them using an IP five-tuple. The values used in the five-tuple may have been signaled during application-layer signaling, for example, Session Initiation Protocol (SIP) [3] signaling in the case of an IP multimedia subsystem (IMS)-voice session. This is described in more detail later. The default bearer typically is not associated with any specific packet filters. Rather, it typically uses a “match all” packet filter, meaning that any packet that does not match any of the existing dedicated bearer packet filters is mapped onto the default bearer. This has the consequence that if a dedicated bearer is dropped by the system, the packet flows that originally were carried on that bearer are rerouted to the default bearer because in that case,

that traffic only matches the “match all” packet filter.

For more details about the EPS bearer, see [4].

QoS PARAMETERS

This section introduces the QoS parameters defined for EPS and explains their purpose and intended use. Additional information about the QoS parameters can be found in [4].

The EPS QoS concept is class-based, where each bearer is assigned one and only one QoS class identifier (QCI) by the network. The QCI is a scalar that is used within the access network as a reference to node-specific parameters that control packet-forwarding treatment (e.g., scheduling weights, admission thresholds, queue management thresholds, link-layer protocol configuration, etc.) and that were preconfigured by the operator owning the node (e.g., the LTE base station).

Each standardized QCI is associated with standardized QCI characteristics. The characteristics describe the packet-forwarding treatment that the bearer traffic receives edge-to-edge between the terminal and the gateway in terms of bearer type (GBR or non-GBR), priority, packet-delay budget, and packet-error-loss rate. See [2] for details. The standardized QCI characteristics are not signaled on an interface. They should be understood as guidelines for the preconfiguration of node-specific parameters for each QCI. The goal of standardizing a QCI with corresponding characteristics is to ensure that applications/services mapped to that QCI receive the same minimum level of QoS in multi-vendor network deployments and in the case of roaming.

Whereas the QCI specifies the user-plane treatment that the packets carried on the associated bearer should receive, the allocation and retention priority (ARP) specifies the control-plane treatment that the bearers receive. More specifically, the ARP enables the EPS system to differentiate the control-plane treatment related to setting up and retaining bearers. That is, the ARP is used to decide whether a bearer establishment or modification request can be accepted or must be rejected due to resource limitations. In addition, the ARP can be used to decide which bearer to release during exceptional resource limitations.

The maximum bit rate (MBR) and GBR are defined only for GBR bearers. These parameters define the MBR, that is, the bit rate that the traffic on the bearer may not exceed, and the GBR, that is, the bit rate that the network guarantees (e.g., through the use of an admission control function) it can sustain for that bearer. In 3GPP Release 8, the MBR must be set equal to the GBR, that is, the guaranteed rate is also the maximum rate that is allowed by the system. Allowing the setting of an MBR greater than a GBR is a candidate for future 3GPP releases. The main scenario that is targeted by such a setting is enhanced support for adaptive video applications where only a minimum video quality is guaranteed by the network.

The main purpose of the aggregate maximum bit rate (AMBR) is to enable operators to limit the total amount of bit rate consumed by a sin-

gle subscriber. As such, it is not defined per bearer, but rather per group of non-GBR bearers. This parameter gives operators the tools to offer differentiated subscriptions that are widespread among operators employing fixed-line broadband technologies (such as digital subscriber line [DSL], e.g., 10 Mb/s or 100 Mb/s download bit rate).

The 3GPP has agreed on defining two different AMBR parameters:

- APN-AMBR: defined per subscriber and APN and known only to the gateway
- terminal-AMBR: defined per subscriber and known by both the gateway and the radio-access network (RAN)

Both of these AMBR values are defined for an aggregate of non-GBR bearers. Bit rate consumed by GBR bearers is not included in either of the AMBR parameters. It should be noted that each of these AMBR values are defined separately for uplink (UL) and downlink (DL) direction; that is, in total, four AMBR values are defined: UL APN-AMBR, DL APN-AMBR, UL terminal-AMBR, and DL terminal-AMBR.

One example of a scenario where this is useful is if an operator offers two separate services: corporate access for virtual private network (VPN) access into corporate networks and Internet access for general access to the Internet. The operator provides these services using separate APNs. The current AMBR definitions enable operators to differentiate the service level provided for each of these services. For example, a subscriber using both of these services could have a 100-Mb/s downlink limit on the corporate access service (i.e., DL APN-AMBR = 100-Mb/s for that subscriber on that APN) and a 5-Mb/s downlink limit on the Internet access service (i.e., DL APN-AMBR = 5 Mb/s for that subscriber on that APN).

A subscribed terminal-AMBR is associated with each subscription. This subscribed value should be considered to be an upper limit of the total bit rate that can be provided to that subscriber. The actual terminal-AMBR that is enforced by the network nodes is then calculated as the minimum of the subscribed terminal-AMBR and the sum of the APN-AMBR of all active APNs (i.e., APNs where the terminal has set up a default EPS bearer).

For a functional view of where the different AMBRs are enforced, see the second subsection in the following section.

QoS MECHANISMS

The mechanisms that are used to provide QoS in the EPS system can be divided into control-plane signaling procedures and user-plane functions, each described in a separate subsection below.

Control-Plane Signaling Procedures — The policy controller in the network determines how each packet flow for each subscriber must be handled in terms of the QoS parameters to be associated with the handling of that packet flow. The policy controller can issue so-called policy and charging control (PCC) rules to the gateway, which in turn are used as a trigger to establish a new bearer or modify an existing bearer to handle a specific packet flow or to modify the han-

dling of a packet flow. The packet flow is described by the UL/DL packet filters. The bearer-level request is forwarded to the LTE RAN and — if admitted by all involved network nodes — to the terminal. A high-level view of the signaling flow is shown in Fig. 3.

The next section provides a brief description and discussion of how QoS control is triggered in the absence of a policy controller (or equivalent node) in the network.

In addition to these dynamic control-plane signaling procedures, the operator must do a semi-static configuration of QoS functions directly in the network nodes through an operation and maintenance (O&M) system. An example of this is the semi-static configuration of node-internal functions (e.g., scheduling functions).

User-Plane Functions — The configuration of the network nodes (both through signaling procedures specified by 3GPP and through an O & M system) enables them to carry out user-plane QoS functions. These functions can be allocated to different nodes and classified into functions that operate per packet flow, per bearer (or group thereof), or per DSCP as illustrated in Fig. 4.

Packet-Flow-Level Functions — 3GPP specifies certain QoS functions that operate on a packet-flow level [2]. Using *packet-flow rate policing*, a gateway (or a physically separate network node) can identify certain packet flows using (*deep*) *packet inspection* techniques [5] and throttle the bit rate experienced by that particular packet flow, without modifying the bearer-level QoS parameters. This can be a useful QoS function for enabling an operator to limit the throughput experienced by a so-called “flat-rate abuser,” that is, a subscriber with a flat-rate pricing plan that engages in extensive uploads or downloads (typically through peer-to-peer applications).

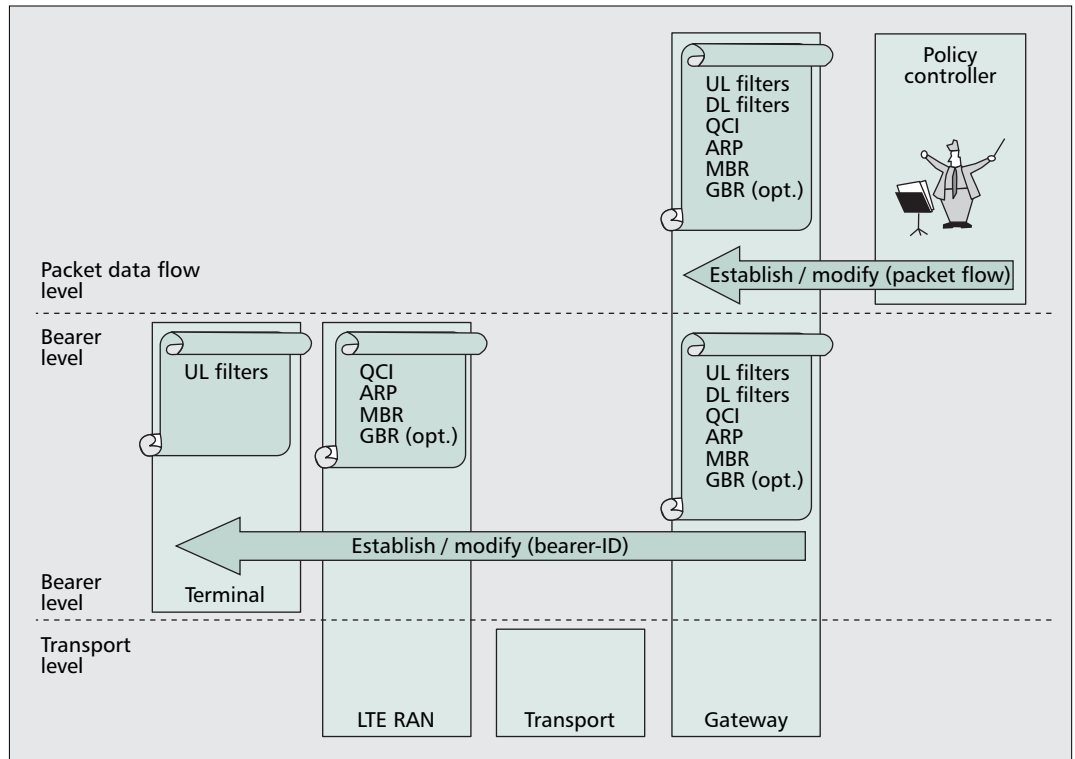
Bearer-Level Functions — The terminal and gateway perform uplink and downlink *packet filtering*, respectively, to map the packet flows onto the intended bearer. These are the underlying functions that provide the network with traffic separation functionality.

The gateway and the LTE RAN can implement functions related to admission control and pre-emption handling (i.e., congestion control) to enable these nodes to limit and control the load put on them. These functions can take the ARP value as an input to differentiate the treatment of different bearers in these functions. For example, the ARP can be used by the pre-emption function to determine which bearers to release from the system in situations when the system is overloaded or when resources must be freed up for other purposes (e.g., an incoming emergency call). In such situations, bearers associated with a low allocation and retention priority are released.

The gateway and the LTE RAN further implement functions related to *rate policing*. The goal of these functions is twofold: to protect the network from becoming overloaded and to ensure that the services are sending data in accordance with the specified maximum bit rates (AMBR and MBR). For the non-GBR bearers,

The main purpose of the aggregate maximum bit rate is to enable operators to limit the total amount of bit rate consumed by a single subscriber. As such, it is not defined per bearer, but rather per group of non-GBR bearers. This parameter gives operators the tools to offer differentiated subscriptions that are widespread among operators employing fixed-line broadband technologies like DSL..

To distribute RAN resources between the established bearers, the LTE RAN implements uplink and downlink scheduling functions. The scheduling function is to a large extent responsible for fulfilling the QoS characteristics associated with the different bearers.



■ Figure 3. A high-level view of EPS signaling procedures to control QoS functions.

the gateway performs rate policing based on the APN-AMBR value(s) for both uplink and downlink traffic, whereas the LTE RAN performs rate policing based on the terminal-AMBR value for both uplink and downlink traffic. For GBR bearers, MBR policing is carried out in the gateway for downlink traffic and in the LTE RAN for uplink traffic.

To distribute RAN resources (radio and processing resources) between the established bearers, the LTE RAN implements *uplink and downlink scheduling* functions. The scheduling function is to a large extent responsible for fulfilling the QoS characteristics associated with the different bearers.

The LTE RAN is responsible for *configuring the L1 and L2 protocols* of the radio connection of the bearer in accordance with the QoS characteristics associated with the bearer. Among others, this includes configuring the error-control protocols (e.g., modulation, coding, and link-layer retransmissions) so that the QoS characteristics packet-delay budget and packet-error loss are fulfilled.

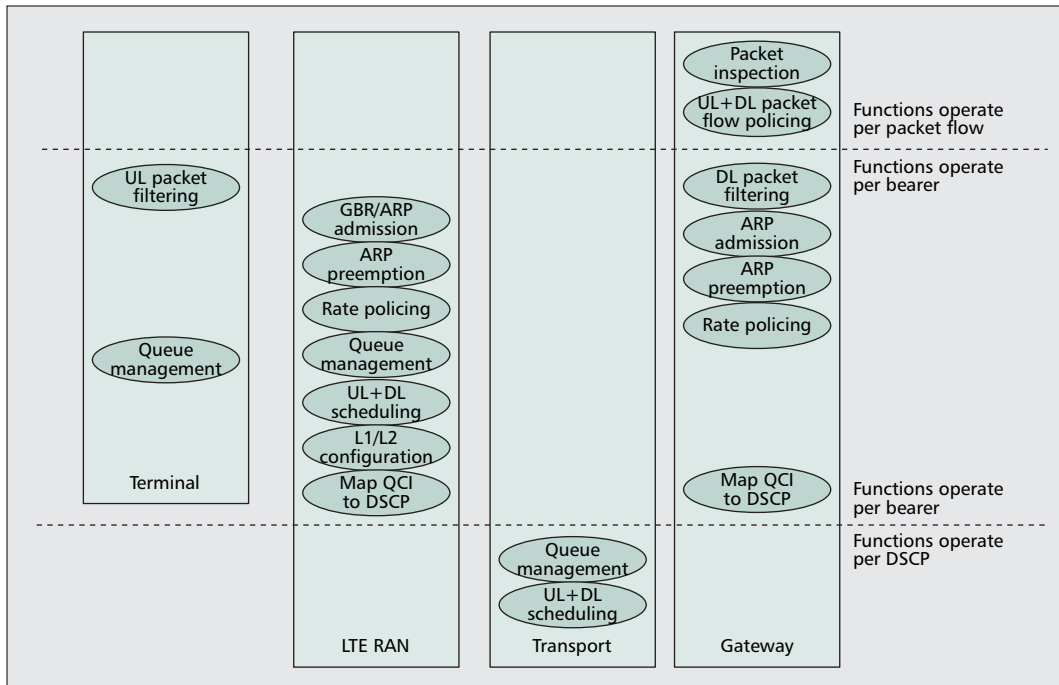
To allow for traffic separation in the transport network, the gateway and the LTE RAN implement a *QCI to DSCP mapping function*. The purpose of this function is to make a translation from bearer-level QoS (QCI) to transport-level QoS (DSCP). Using this function, packets on a bearer associated with a specific QCI are marked with a specific DSCP for forwarding in the transport network. The QCI to DSCP mapping is performed based on operator policies. These are configured into the network nodes through an O & M system. For downlink packets, the gateway performs this mapping while the LTE RAN performs it for uplink packets.

DSCP-Level Functions — Transport network nodes can implement *queue management* schemes and scheduling algorithms for uplink and downlink traffic. In the transport network, the bearer is not visible; and hence, these algorithms determine the traffic forwarding treatment of each individual packet, based on the DSCP value.

NETWORK- AND TERMINAL-INITIATED QoS CONTROL

There are two different paradigms that can be used to establish a dedicated bearer with a specific QoS in EPS. We refer to these as the terminal-initiated and network-initiated QoS control paradigms. The background and motivation for introducing a network-initiated paradigm into the 3GPP specifications was originally described in [1]. This paradigm subsequently was introduced into both the general packet-radio service (GPRS) 3GPP Release 7 specifications [6] (covering 2G/3G accesses), the EPS 3GPP Release 8 specifications (covering system architecture evolution [SAE]/LTE) [4], as well as into the evolved high-rate packet data (eHRPD) system specified in 3GPP2 [7]. The basic principles are shown in Fig. 5.

Using network-initiated QoS control, the network initiates the signal to set up a dedicated bearer with a specific QoS toward the terminal and the RAN. This is triggered by an application function (AF) or a deep-packet inspection (DPI) function [2, 5, 8], and the signal is carried over standardized interfaces (Rx and/or Gx). Using this paradigm, the client application can be left “access QoS unaware,” meaning that it is not



■ **Figure 4.** Overview of user-plane QoS functions in EPS.

The main motivation for specifying the network-initiated QoS-control paradigm is that services are typically provided by the access network operator. As such, it is natural that the access network and service owner assigns the QoS level per packet flow associated with a particular service.

required to be aware of the specifics of the QoS model of the access network. However, typically, the client application has access-agnostic knowledge of the QoS with which it wants to be provided. For some services, the QoS to be applied to the session can be negotiated with the network by means of application-layer signaling, such as SIP [3] and Real Time Streaming Protocol (RTSP) [9]. It is important to note, however, that there is no access-specific information in this signaling.

Using a terminal-initiated QoS control paradigm, it is the terminal that initiates the signal to set up a dedicated bearer with a specific QoS toward the network (which in turn triggers a command to the RAN). The trigger for this signal is carried over a terminal vendor-specific QoS application programming interface (API). This means that to specify the QoS information for the bearer, the client application must be “access QoS aware,” meaning that it must be aware of the specifics of the QoS model of the access network. In this case, there is no policy controller communicating any QoS information to the network.

The main motivation for specifying the network-initiated QoS-control paradigm is that services (e.g., Internet access, mobile-TV, IMS voice) are typically provided by the access network operator, potentially through peering agreements with third-party service operators. As such, it is natural that the access network and service owner assigns the QoS level per packet flow associated with a particular service.

Network-initiated QoS control minimizes the terminal involvement in QoS and policy control. It has the following key advantages when compared to terminal-initiated QoS control:

- It can be used to provide QoS to access-agnostic client applications, such as applications that are downloaded and installed by

the subscriber. This is not possible for terminal-initiated QoS control, which requires access-specific client applications that must be programmed toward a QoS API that is specific to the terminal vendor.

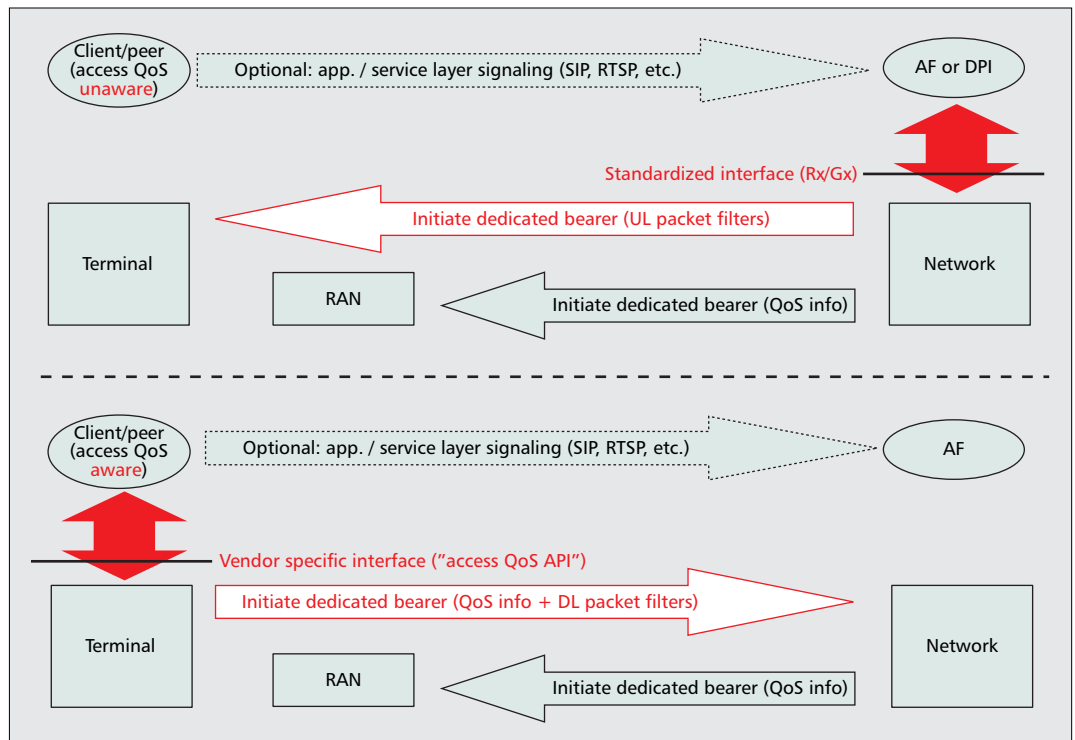
- As a direct result of the previous, network-initiated QoS control enables QoS to be provided in the “split-terminal” case where the client applications resides in a node (e.g., a laptop or set top box) that is physically separated from the terminal.
- It enables the deployment of more consistent exception-handling policies. One example of such a policy is the specification of the action to take when the request to initiate a service (or associated bearer) is rejected. There are numerous possible actions to take (e.g., give up, retry N times, or retry with a lower QoS-level). This consistency can be achieved more easily because using the network-initiated QoS control paradigm, the policies are centralized in the network rather than distributed in the numerous terminals from multiple vendors, as is the case using the terminal-initiated QoS control paradigm.

Due to the advantages listed above, we regard the network-initiated QoS control paradigm to be the most useful in cases where the operator controls the service (see [10] for a definition of operator-controlled services). For non-operator-controlled services, there is also the possibility to use the terminal-initiated QoS control paradigm. However, this possibility is not elaborated in this article.

END-TO-END USE CASE

In this section, we present an end-to-end use case in which a subscriber sets up an IMS voice call where the EPS QoS concept is used to real-

When the media flow starts, the packet filters in the terminal and gateway map the IMS voice-over-IP (VoIP) packets onto the dedicated bearer, and the IMS VoIP service for this subscriber receives the QoS treatment defined by its service and subscriber differentiation policies.



■ **Figure 5.** Illustration of differences in the network-initiated (top) and terminal-initiated (bottom) QoS control paradigms.

ize the QoS. In particular, the network-initiated QoS control paradigm as described in the previous section is applied. The intent of this use case is to illustrate how an operator can make use of the described mechanisms for providing subscriber and service differentiation.

In addition to the terminal, LTE RAN, transport network, and gateway, the system consists of a policy controller and an application function. The latter is a call-state control function (CSCF) in the IMS architecture [11]. The system and the signaling in the use case are illustrated in Fig. 6.

At the start of the use case, the subscriber is engaging in two services, Internet browsing and peer-to-peer file sharing. These services are both mapped onto the default bearer (shown in grey). The IMS application in the client was preconfigured with the IP address of the CSCF so that signaling messages are directed toward this node.

The subscriber places an IMS voice call, and the media flow is preceded by application-layer signaling using the SIP protocol to set up the call (1). This end-to-end signaling is intercepted by the CSCF in the network, and the messages reveal the IP five-tuple, as well as access-agnostic QoS information (see description in previous section) to the CSCF. One example of this information is codec rates. Based on this information, the CSCF detects a new packet flow and passes this information to the policy controller (2). The policy controller uses the information provided by the CSCF, operator-defined service policies, and subscription data when determining the appropriate QoS treatment that the packet flow should receive. This treatment is signaled to the gateway through the QoS parameters, and the

packet flow is described in the defined uplink and downlink packet filters (3). At the reception of this information, the gateway initiates a dedicated bearer-establishment procedure in the control plane. This procedure sets up the dedicated bearer (4) and configures the user-plane QoS functions so that the packets carried on that bearer receive the appropriate QoS treatment. When the media flow starts, the packet filters in the terminal and gateway map the IMS voice-over-IP (VoIP) packets onto the dedicated bearer, and the IMS VoIP service for this subscriber receives the QoS treatment defined by its service and subscriber differentiation policies.

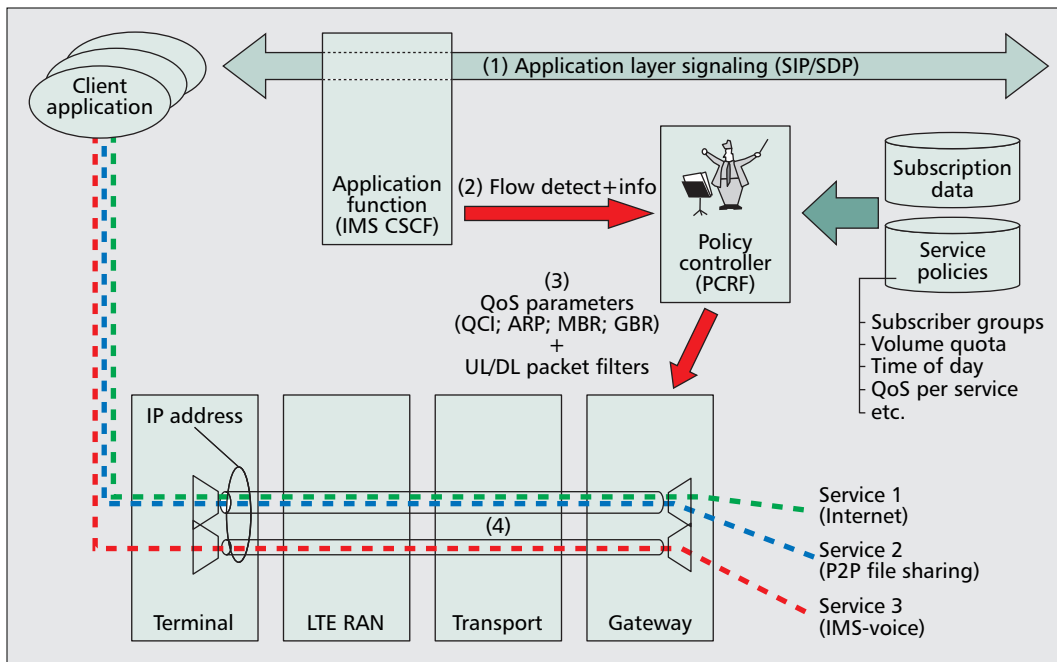
SUMMARY AND CONCLUSIONS

In this article, we described the QoS concept of the EPS that was standardized in the 3GPP Release 8 specifications. This concept is based on two fundamental principles:

- Network-initiated QoS control
- Class-based mapping of operator services to packet-forwarding treatment in user-plane nodes

The driver for introducing both principles was to simplify and enhance operator control over the provisioning of services and their associated QoS. This is achieved with the evolved QoS concept because it minimizes terminal involvement in QoS and policy control and centralizes the execution of operator policies in the network.

With the network-initiated QoS control paradigm, only the network can make the decision to establish or modify a bearer. This is a shift from the terminal-initiated QoS control paradigm in pre-Release 7, where this decision



■ Figure 6. Illustration of a use case where a subscriber sets up an IMS voice call with the EPS QoS concept.

The goal of standardizing a QCI with corresponding characteristics is to ensure that applications or services that are mapped to that QCI receive the same minimum level of QoS in multivendor network deployments and in the case of roaming.

can be made only by the terminal. Network-initiated QoS control has a number of advantages: it can be used to provide QoS to access-agnostic client applications (such as those downloaded and installed by the subscriber), it enables QoS to be provided in the split-terminal case, where the client application resides in a node (e.g., a laptop or set top box) that is physically separated from the terminal, and finally, it enables the deployment of more consistent exception-handling policies.

This paradigm assumes that there is network intelligence, for example, application functions or deep-packet inspection functions, that can both identify the service that a subscriber is initiating and trigger QoS control (e.g., setting up a new bearer) when required.

The class-based approach to mapping of operator services to packet-forwarding treatment is a shift from the flow-based approach specified in 3GPP Release 7. With the class-based approach, an operator maps supported applications or services to a small set of QoS classes. Thereby, each packet flow is associated with one and only one QCI. The QCI is a scalar that is used as a reference to node-specific parameters that control packet-forwarding treatment and that are preconfigured by the operator owning the user-plane node. The 3GPP Release 8 specifications include nine standardized QCIs with corresponding standardized characteristics in terms of bearer type (GBR versus non-GBR), priority, packet delay, and packet-error-loss rate. The goal of standardizing a QCI with corresponding characteristics is to ensure that applications or services that are mapped to that QCI receive the same minimum level of QoS in multivendor network deployments and in the case of roaming.

The combination of these two fundamental principles, network-initiated QoS control and class-based mapping of services, provides access-

network operators and service operators with a set of tools to enable service and subscriber differentiation. These tools are becoming increasingly important as operators are moving from a single to a multi-service offering at the same time as both the number of mobile broadband subscribers and the traffic volume per subscriber is increasing rapidly.

REFERENCES

- [1] R. Ludwig *et al.*, "An Evolved 3GPP QoS Concept," *Proc. IEEE VTC*, Spring 2006.
- [2] 3GPP Tech. Spec. 23.203, "Policy and Charging Control Architecture," v. 8.3.1.
- [3] J. Rosenberg *et al.*, "SIP: Session Initiation Protocol," RFC 3261, www.ietf.org, June 2002.
- [4] 3GPP Tech. Spec. 23.401, "General Packet Radio Service (GPRS) Enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) Access," v. 8.3.0.
- [5] "Traffic Inspection for Visibility, Control and New Business Opportunities," White Paper, Feb. 2008, <http://www.ericsson.com>
- [6] 3GPP Tech. Spec. 23.060, "General Packet Radio Service (GPRS): Service Description," v. 7.8.0.
- [7] 3GPP2 Tech. Spec. X.P0057-0, "E-UTRAN — eHRPD Connectivity and Interworking: Core Network Aspects," v. 0.13.0.
- [8] J. J. Pastor, S. Rommer, and J. Stenfelt, "Policy and Charging Control in the Evolved Packet System," *IEEE Commun. Mag.*, Special Issue on Next Generation 3GPP Technologies, Feb. 2009.
- [9] H. Schulzrinne *et al.*, "Real Time Streaming Protocol (RTSP)," RFC 2326, <http://www.ietf.org>, Apr. 1998.
- [10] 3GPP Tech. Rep. 23.882, "3GPP System Architecture Evolution: Report on Technical Options and Conclusions," v. 8.0.0.
- [11] 3GPP Tech. Spec. 23.228, "IP Multimedia Subsystem (IMS)," v. 8.6.0.

BIOGRAPHIES

HANNES EKSTRÖM (hannes.ekstrom@ericsson.com) joined Ericsson in 2000. He is currently a product manager with responsibility for QoS functionality in the Ericsson packet core product portfolio. He has held various positions within Ericsson including at Ericsson Research, where he worked in performance evaluation and concept development of wireless networks, and at LTE Systems Management, where he developed radio resource management algorithms.