An Evolved 3GPP QoS Concept

Reiner Ludwig, Hannes Ekström, Per Willars and Niklas Lundin Ericsson Research & Development

Abstract- Simple and cost efficient means for providing and controlling Quality of Service (QoS) are important for 3GPP operators. This is particularly important for operators planning to provide a wide range of IP-based services across 3GPP broadband shared channels. We point out shortcomings of today's (Rel. 5) 3GPP QoS concept and, based on these, explain why that demand is currently not met. Based on a set of requirements identified for an evolved 3GPP QoS concept, we propose four small additions to the 3GPP specifications. The resulting evolved QoS concept is a realization of DiffServ for 3GPP access networks enhanced with the integration with session admission control. Operator QoS control is exercised from the Policy Charging Rule Function (PCRF) and through preconfiguration via the management plane. The evolved QoS concept provides a 3GPP operator with capabilities beyond those found in state-of-the-art fixed broadband access networks.

Keywords- 3G Evolution, Quality of Service (QoS), DiffServ

I. INTRODUCTION

Simple and cost efficient means for providing and controlling Quality of Service (QoS) are important for access network operators. However, we recognize that today's (Rel. 5) QoS concept of the 3rd Generation Partnership Project (3GPP) does not meet that demand. This becomes particularly evident for operators planning to provide a range of IP-based [5] services across 3GPP broadband shared channels.

The goal of this paper is (1) to establish a set of requirements for an evolved 3GPP QoS concept, and (2) to identify possible additions to the 3GPP specifications as a proposal for how the 3GPP Rel. 5 QoS concept can be evolved to meet those requirements. Our proposal adopts the realization of a number of elements from state-of-the-art fixed broadband access networks such as

- network-controlled bearer handling (e.g., see [1], [2]),
- network-controlled uplink and downlink packet classification (e.g., see [1], [8]), and
- class-based traffic separation and traffic treatment (e.g., see [1], [6]).

In Section II, we review the key elements that need to be present in an access network to provide QoS. In Section III, we suggest a set of requirements that we believe should be met by an evolved 3GPP QoS concept. In Section IV, we review how the mentioned elements are realized in the 3GPP Rel. 5 QoS concept, and point out limitations with respect to the fulfillment of the requirements. In Section V, we use Digital Subscriber Line (DSL) access to briefly explain how the elements are realized in a fixed broadband access network. In Section VI, we present the four proposed additions to the 3GPP specifications. In Section VII, we summarize and conclude the paper.

II. KEY ELEMENTS FOR PROVIDING ACCESS QOS

In this section, we review the key elements that need to be present in an access network to provide QoS, and introduce the terminology used throughout this paper (see Fig. 1). It has been our aim to come up with generic terminology that should be applicable to a wide range of access technologies, i.e., we have not limited ourselves to the terminology defined in existing 3GPP specifications. The User Equipment (UE) is the terminal device, the Radio Access Network (RAN) includes the base station and other radio-related infrastructure, and the Gateway (GW) provides the IP connectivity.

We use the term service as the offering an operator makes to a subscriber. Examples of a service include VoIP telephony based on the IP Multimedia Subsystem (IMS), Mobile-Television, Internet-Access (with various levels of user differentiation), Instant Messaging, MBMS (Multimedia Broadcast Multicast Service) and PoC (Push-to-Talk over Cellular). We further distinguish between session-based services and non-session-based services. Session-based services utilize an end-to-end session control protocol such as SIP/SDP or RTSP/SDP [8], [10], [11]. All IMS services are session-based while Internet-Access is an example of a nonsession-based service. The traffic running between a particular client application and a service can be differentiated into separate service data flows. For example, an IMS-VoIP session can be differentiated into two service data flows, one for the session control signaling, and one for the media.

We use the term *Traffic Forwarding Policy (TFP)* to denote a set of pre-configured traffic handling attributes relevant within a particular user plane network element. For example, a RAN-TFP may include several attributes such as the link layer protocol mode (acknowledged or unacknowledged), the power



Figure 1. Key Elements for Providing QoS in an Access Network

0-7803-9392-9/06/\$20.00 (c) 2006 IEEE

settings, and a default uplink maximum bit rate; while a GW-TFP may only include a default downlink maximum bit rate. Each edge/bottleneck node – potentially including transport network nodes – supports a number of TFPs. Uplink (UL) and Downlink (DL) Guaranteed Bit Rates (GBRs) are not part of a TFP since these traffic handling attributes can not be preconfigured for a QoS class. They must therefore instead be dynamically signaled. TFPs confine traffic handling attributes to those nodes where those attributes are actually needed. TFPs are provided and configurable by the operator from the management plane.

We use the term *bearer* to refer to an edge-to-edge association between the UE and the GW. Independent of whether it is realized in a connection-oriented or a connectionless way, a bearer is defined through

- 1. the network to which it connects the UE (referred to as Access Point Name in 3GPP),
- 2. the *QoS Class Identifier (QCI)* via which it can be associated with a TFP defined within each user plane edge/bottleneck node, and
- 3. (optionally) the UL- and DL-GBR. Within an access network the UL-GBR and DL-GBR are only relevant for session-based services, and only if the operator's policy defined for a specific QoS class requires that session admission control (e.g., in the RAN) be triggered when establishing service data flows associated with that QoS class.

Note that the term QCI is not associated with any semantics, e.g., related to traffic characteristics or applicationlayer requirements on end-to-end QoS. That is, a QCI is simply a "pointer" to a TFP. Note further that within a specific node multiple QCIs may be associated with the same TFP.

In order to receive a QoS level other than the default QoS level (via the default bearer explained below) a service data flow needs to be bound to what we refer to as a QoS bearer. A QoS bearer is associated with an *uplink binding state* in the UE for the uplink traffic, and a *downlink binding state* in the GW for the downlink traffic. The binding state creates the mapping of a service data flow to a QoS bearer. When multiple bearers are established between a UE and a GW then it is the uplink binding states of the QoS bearers that "steer" the aggregate uplink traffic into the right bearers, and likewise the downlink binding states in the GW for the aggregate downlink traffic. For a specific network that the UE connects to there is at most one bearer without uplink and downlink binding states. This bearer is referred to as the *default bearer*. It is important that each QoS bearer connecting to a specific network is associated with well defined ("non-overlapping") uplink and downlink binding states to ensure an unambiguous mapping of packets to the QoS bearers.

As we show in Section IV, all the elements described above are present in the 3GPP Rel. 5 QoS concept. In Section V, we further show that they are also present in state-of-the-art fixed broadband networks (e.g., DSL). The main differences between both access technologies lies in the control procedures (terminal-based vs. network-based control), and the representation of the information elements.

III. REQUIREMENTS FOR AN EVOLVED 3GPP QOS CONCEPT

Below we list requirements that we believe should be met by an evolved 3GPP QoS concept followed by a brief discussion of each requirement.

- a) Operator Controlled Service and User Differentiation
- b) Minimize Terminal Involvement in QoS and Policy Control
- c) QoS Support for Access Agnostic Client Applications (UE-Based + Non-UE-Based)
- d) Fast Session Setup
- e) Backwards Compatibility
- f) Convergence towards Other Access Types such as I-WLAN [14] and fixed broadband [4]
- g) Rapid Time To Market (TTM) for the Deployment of New Services

Requirement a)

Service and user differentiation requires a limited set of welldefined QoS classes. The number of QoS classes supported within an operator's network reflects the granularity of differentiation the operator provides. Operators should be free to define the mapping of the service data flow(s) of offered services to the QCI(s). For certain well-known services this mapping could be standardized, or defined as part of roaming agreements. Likewise, operators should be free to define which TFP gets associated with a QCI.

Requirement b)

Operators may regard a UE as a non-trusted device which can be "hacked", e.g., for the purpose of receiving higher QoS than subscribed and charged for. Therefore, the control over a bearer's QCI should be located within the network. In principle, there is no reason for a UE to have knowledge of a bearer's QCI. Another aspect of this requirement is the placement of the exception handling control associated with bearer establishment. To ensure a consistent exception handling across terminals from different vendors, this control should be located within the network.

Requirement c)

Access agnostic client applications do not use any vendorand/or access-specific QoS-API (Application Programming Interface). A QoS-API can be used to request the establishment of a QoS bearer, and thereby create the UL binding between a service data flow of the requesting client application and the QoS bearer. This requirement basically says that any client application programmed towards the ubiquitous socket-API that is supported by virtually every widely deployed operating system should be able to receive QoS. Note that the socket-API does not support requests for QoS bearers.

Requirement d)

It is widely recognized that low session setup delays are an important factor in user perceived service quality.

Requirement e)

It can be expected that UEs based on the 3GPP Rel. 5 QoS concept will be widely deployed in the coming years. Also, the upgrade of network equipment can not be assumed to be carried out "over night". Hence, backwards compatibility with Rel. 5 based equipment needs to be ensured by an evolved 3GPP QoS concept.

Requirement f)

An evolved 3GPP QoS concept should be aligned with the QoS concepts of other access types such as I-WLAN and fixed broadband networks. This will facilitate and simplify the provisioning of end-to-end QoS in multi-access networks.



Figure 2. Philosophy behind the 3GPP Rel. 5 QoS Concept

Requirement g)

An evolved 3GPP QoS concept should allow for deployment of operator-defined services without the need for prior standardization of QoS support in UEs or network elements.

IV. REVIEW OF THE 3GPP REL. 5 QOS CONCEPT

In this section, we briefly review how the elements outlined in Section II are realized in the 3GPP Rel. 5 QoS concept, and evaluate to what extent the requirements set forth in the previous section are met.

As depicted in Fig. 2, the philosophy behind the 3GPP Rel. 5 QoS concept is based on the assumption that the information about the requested service (e.g., application-layer QoS requirements) is only present in the UE, and that this information must be provided to the network from the requesting client application.

It must be said, though, that when this concept was devised more than 10 years ago, this was the only reasonable assumption that could be made. At that time, the Internet had not yet emerged to become what it is today. A network-based service logic for packet-based services such as IMS did not exist, IP-based session control protocols such as SIP/SDP and RTSP/SDP had not been standardized, and likely only few people had thought of network-based packet inspection mechanisms that would allow the network to infer the requested service.

This has led to a number of important consequences which we today perceive as limitations:

- A QoS bearer can only be initiated from the UE.
- The uplink and downlink binding states are controlled from the UE.
- The QCI is represented as a record of attributes referred to as 3GPP QoS profile [12].

One limitation resulting from UE-initiated establishment of QoS bearers is that it leaves the exception handling to the UE's local policy. For example, this policy could be to retry setting up the QoS bearer with the same or different Requested QoS, or perhaps to simply give up trying to set up the call. This is in conflict with requirement b). Furthermore, the absence of the possibility to initiate the establishment of a QoS bearer from the network (using a bearer handling procedure) precludes the possibility to pre-activate QoS bearers based on operator policy. This is in conflict with requirement d). We believe that an important step towards reducing setup delays is the exploration of solutions for the pre-activation of QoS bearers.

With the 3GPP Rel. 5 QoS concept a vendor-specific QoS-API has to be used by application developers to explicitly bind a service data flow from a UE-based client application to a specific QoS bearer in the uplink direction, i.e., to create the UL binding state. This only works for UE-based client applications. Non-UE-based client applications hosted, e.g., on a general-purpose computing device connected to the UE, can only receive default QoS. This limitation is in conflict with requirement c), and can be perceived as a barrier to third party mobile client application development, and thus indirectly to the uptake of mobile services.

The 3GPP Rel. 5 QoS concept does not meet requirement a) for a number of reasons. The mentioned 3GPP QoS profile is signaled between UE, RAN, and GW during bearer establishment. The idea behind the QoS profile is that it should include all the traffic handling attributes required by each user plane network element on the path to control the traffic of the associated bearer. A consequence of this approach is that some attributes have to be stored in nodes where they are never used. In practice, however, the QoS profile is merely used as a pointer, i.e., similar to a QCI, to identify traffic handling attributes that have not been specified as part of the QoS profile. For example, within the RAN a QoS profile points to a so-called radio bearer configuration [13] where the only attributes actually used from the QoS profile is the Traffic Class (conversational, streaming, interactive, and background) and the UL- and DL-GBR. One problem that is emerging as an increasing number of IP-based services are being offered is the lack of a sufficient number of well-defined QCIs. When treating the GBR attributes separately, only four different QCIs can be defined (conversational, streaming, interactive, and background).

One approach to "generate" more QCIs is to use other attributes in addition such as the 3 Traffic Handling Priorities defined for the Traffic Class 'interactive'. This still only provides six QCIs of which only two can be used with session admission control since UL- and DL-GBR are only defined for the Traffic Classes 'conversational' and 'streaming'. One could continue this approach of "generating" more QCIs by also using other attributes. But, it may have unforeseen side-effects in already deployed equipment if attributes are suddenly used when they were previously ignored.

Another issue with using the 3GPP QoS profile (without UL- and DL-GBR) as a QCI is that it is associated with traffic requirements. This is because the attribute Traffic Class not only identifies a QoS class but in addition expresses the delay sensitivity of the service data flow mapped onto the corresponding bearer. This is actually a semantic overload of this attribute. Recall from Section II, that a QCI is simply a pointer that is otherwise not associated with any semantics related to traffic characteristics or application-layer requirements on end-to-end QoS. This may lead to debate and/or misunderstandings, e.g., if a best-effort VoIP service were mapped to the Traffic Class 'background'.

V. COMPARISON WITH FIXED BROADBAND ACCESS

In this section, we use DSL access to briefly explain how the elements outlined in Section II are realized in a fixed broadband access network. We also show how the requirements set forth in Section III are met.

QoS bearers in DSL access [1] are realized as connectionless Ethernet routes between the Customer Premises Equipment (CPE), the Digital Subscriber Line Access Multiplexer (DSLAM), and the Broadband Remote Access Server (BRAS) which correspond to a UE, RAN and a GW, respectively. As QCI so-called 802.1Q VLAN tags [3] are used on Ethernet level, and DiffServ Code Points (DSCPs) [7] on IP level. TFPs and QCI→TFP mappings are configured into the DSLAMs and BRAS from the management plane. DL binding state (aka traffic classification state) is located in the BRAS and controlled from the network. UL binding state is also

390

controlled from the network using the Dynamic Host Configuration Protocol (DHCP) with an optional extension [8] to install a static route for each QoS bearer either directly into the (routed) CPE, or into the device (e.g., a personal computer) connecting to the (bridged) CPE. A static route is simply an uplink packet filter that in this case only filters on the destination IP address. On the DSL line between CPE and DSLAM an Ethernet VLAN is today usually realized as a separate Asynchronous Transfer Mode (ATM) Permanent Virtual Circuit (PVC). The establishment and modification of the ATM PVCs is controlled from the network using the Integrated Local Management Interface (ILMI) protocol [2]. Efforts have been started to further enhance fixed broadband access networks, e.g., through the introduction of standardized interfaces required for session admission control and policy and charging control functions [4].

All the requirements set forth in Section III are met by this concept: a sufficient number of QCIs is available for an operator, and the QCI itself is just a numeric value, i.e., simply a pointer. The involvement of the CPE is reduced to a minimum. The only action performed by the CPE is uplink packet classification which is controlled from the network by means of the static routes installed via DHCP. Gating control in the BRAS verifies that the CPE is well-behaved, i.e., that it only maps packets to a static route that have a matching destination IP address. The CPE does not have access to a bearer's QCI since the uplink packet marking is performed in the DSLAM. Bearer establishment and modification can not be controlled from the CPE, but only from the network. To avoid setup delays QoS bearers in DSL are usually pre-established and can remain established for days and weeks.

VI. PROPOSED ADDITIONS TO THE 3GPP QOS CONCEPT

We propose the following four small additions to the 3GPP QoS concept:

- 1. Use of Session-Level Signaled data to Identify a Requested Session-Based Service
- 2. A New Network-Initiated QoS Bearer Establishment Procedure
- 3. A New Network-Controlled Procedure to Install Uplink Binding State in the UE
- 4. A New QoS Class Identifier Between GW and RAN

The proposed additions are independent of each other, although it seems natural to realize 2. and 3. in a combined procedure.

We believe that an evolved 3GPP QoS concept should also support the philosophy that the information about the requested service (e.g., application-layer QoS requirements) can be inferred by the network. Based on that information combined with an operator's policies, the network can create the required UL binding state, and initiate the establishment of a QoS bearer. This approach is depicted in Fig. 3.

There are at least three different ways the network can infer the required information about a requested service. For nonsession-based services such as (standard or premium) Internet-Access the required information is stored in subscriber profile databases. A second option is to use packet inspection mechanisms in the GW, e.g., to redirect premium content received on the bearer for Internet-Access to a separate and higher priority bearer, or to derive from the RTSP/SDP signaling the DL-GBR required for the streaming media. The introduction of IMS enables a third (from a purist's point



Figure 3. Adding an Alternative Philosophy to the 3GPP QoS Concept

of view, a cleaner) possibility to provide service information to the network. The so-called Application Function (AF) of an IMS service can communicate with the network via the Policy Charging Rule Function (PCRF) across open and standardized interfaces. The AF is involved in the SIP/SDP-based session control signaling running between the involved peering client applications. By forwarding data present in the session signaling such as communication service identifiers and application references to the PCRF, the PCRF is able to identify exactly which service is requested. Together with information about the corresponding service data flows (from SDP), the PCRF would then be able to map each service data flow to a QCI according to operator policy.

A network-initiated QoS bearer establishment procedure allows an operator to ensure a consistent exception handling across terminals from different vendors. In addition, such a procedure enables the possibility of pre-activating QoS bearers based on operator policy, e.g., to reduce setup delays. It should be noted that for pre-activated bearers, we propose to use the possibility to set up a bearer with no resources reserved. If admission control needs to be triggered, this can be done later for an already established bearer. This possibility can be used also when a service is mapped to an already established bearer, which carries traffic for another service.

Furthermore, if after a transition phase backwards compatibility is no longer a concern, and QoS bearers are always established with a network-initiated procedure then this would make the 3GPP QoS profile stored in the UE obsolete. Note that for non-session-based services (e.g., Internet-Access, Messaging, etc.) a UE-based QoS profile is of little use and for session-based services the client application receives the required QoS attributes via the end-to-end session control signaling. Further, other access technologies (e.g., I-WLAN) do not maintain a QoS profile in the terminal.

UE-based UL packet filters per QoS bearer are a possibility to represent the UL binding state in a way that is transparent to the client application. An UL packet filter could potentially filter on the entire 5 tuple of an IP packet (source/destination IP address, source/destination port number, and protocol identifier) and possibly also the DSCP. The set of UL packet filters associated with all established QoS bearers for a UE basically represent a UE-based routing table. This approach makes it possible to bind a service data flow of an access agnostic client application (UE-based + Non-UE-based) to a QoS bearer in the uplink direction.

Note that Non-UE-based client applications are not capable of directly triggering bearer handling procedures in a 3GPP UE. A "copy" of each QoS bearer's UL packet filter should be maintained in the network for gating control purposes, i.e., to verify that the UE maps the "right" packets to a certain QoS bearer.

The installation of UL packet filters into a UE should be coupled with the network-initiated QoS bearer establishment procedure mentioned above. However, also network-initiated modifications of an uplink packet filter



Figure 4. Operator QoS Control

should be supported, e.g., to account for situations where information required to create an UL packet filter is not available at the time when the QoS bearer is established, but only becomes available later during session control signaling (SIP/SDP or RTSP/SDP).

One way to address requirement a) and the shortcomings of the 3GPP QoS profile is to introduce a new information element between the GW and the RAN called QCI. The QCI would simply be a numerical value, and a limited set of QCIs would need to be specified. The idea is to use the QCI as outlined in Section II. That is, within the RAN, the GW, and potential bottleneck nodes of the transport networks the QCI is used as a pointer to a pre-configured TFP. Note that the QCI does not need to be available in the UE. To ensure backwards compatibility during a transition phase the QCI should be signaled "in parallel" between the GW and the RAN as an alternative to the 3GPP QoS profile. That way, nodes that do not support this new information element could fallback to using the 3GPP QoS profile.

Since a QoS bearer's traffic handling attributes are isolated into RAN-TFPs, GW-TFPs, and TFPs of transport network nodes, and assuming an operator can configure at least a subset of the respective traffic handling attributes (e.g., the scheduling priority) an operator can "upgrade" or "downgrade" the QoS of a bearer in a way that is transparent to the UE.

Fig. 4 shows a procedural view of an evolved 3GPP QoS concept including the additions proposed in this section. It should be noted that the way the signaling procedures ("Session Authorization", "PCC Rule", and "Network-Initiated Bearer Activate / Modify") have been depicted should not imply any temporal dependencies. For example, various forms of pre-activation of the bearer are conceivable before a "Session Authorization" takes place. The "Session Authorization" is depicted with a dashed line to indicate that dynamic session authorization is only required for sessionbased services. The UL and DL GBR values are optionally signaled, when the operator has decided to apply admission control for a specific QoS class.

VII. CONCLUSION

In this paper, we have presented a set of requirements that we believe should be met by an evolved 3GPP QoS concept. Based on the requirements we have pointed out shortcomings of the 3GPP Rel. 5 QoS concept. We have then proposed four small and independent additions to the 3GPP specifications:

- 1. Use of Session-Level Signaled data to Identify a Requested Session-Based Service
- 2. A New Network-Initiated QoS Bearer Establishment Procedure
- 3. A New Network-Controlled Procedure to Install Uplink Binding State in the UE
- 4. A New QoS Class Identifier Between GW and RAN

The latter three additions have been adopted from state-ofthe-art fixed broadband access networks. In fact, the resulting evolved QoS concept is a realization of DiffServ for 3GPP access networks enhanced with the integration of session admission control. Operator QoS control is exercised from the PCRF and through pre-configuration via the management plane. The evolved QoS concept provides an operator with capabilities beyond those found in state-of-the-art fixed broadband access networks, e.g., with more advanced uplink packet filtering capabilities. With such an evolved QoS concept operators are well prepared as they move into the multi-service mobile broadband era.

ACKNOWLEDGMENT

This work has greatly benefited from numerous fruitful discussions with many of our colleagues at Ericsson.

REFERENCES

- TR-059, "DSL Evolution Architecture Requirements for the Support of QoS-Enabled IP Services", Technical Report, DSL Forum, September 2003
- [2] ILMI-V4, "Integrated Local Management Interface (ILMI) Specification Version 4.0", ATM Forum, September, 1996
- [3] IEEE Std 802.1Q-1998, "Virtual Bridged Local Area Networks", LAN MAN Standards Committee of the IEEE Computer Society, December 1998
- [4] ES 282 003, "TISPAN Resource and Admission Control Subsystem (RACS) Functional Architecture", ETSI TISPAN, February 2006
- [5] J. Postel, "Internet Protocol", RFC 791, September 1981
- [6] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, "An Architecture for Differentiated Service", RFC 2475, December 1998
- [7] K. Nichols, S. Blake, F. Baker, D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC2474, December 1998
- [8] T. Lemon, S. Cheshire, B. Volz, "The Classless Static Route Option for Dynamic Host Configuration Protocol (DHCP) version 4", RFC3442, December 2002
- [9] J. Rosenberg, et al., "SIP: Session Initiation Protocol", RFC 3261, June 2002
- [10] H. Schulzrinne, A. Rao, R. Lanphier, "Real Time Streaming Protocol (RTSP)", RFC 2326, April 1998
- [11] M. Handley, V. Jacobson, "SDP: Session Description Protocol", RFC 2327, April 1998
- [12] 3GPP TS 23.107, "Quality of Service (QoS) concept and architecture", V6.3.0, June 2005
- [13] 3GPP TS 34.108, "Common test environments for User Equipment (UE)", V6.1.1, January 2006
- [14] 3GPP TR 23.836, "Quality of Service (QoS) and policy aspects of 3GPP - Wirless Local Area Network (WLAN) interworking", V1.0.0, December 2005