

# QoS Control for NGN: A Survey of Techniques

Changho Yun and Harry Perros

**Abstract**—This article presents an overview of standards and a survey of recent work on QoS control of next generation networks (NGNs). We outline primarily standardized QoS control architectures with respect to the service and transport strata of NGN. In the service stratum, QoS parameter negotiation using core IMS procedures is described. In the transport stratum, the resource control and admission schemes of IMS (3GPP), RACF (ITU-T), and RACS (ETSI/ TISPAN) are explained and compared with each other in terms of QoS coverage and number of required signaling messages. In addition, we give a survey of papers on QoS control architectures, which further strengthen existing standards by adaptively managing QoS modification according to user's QoS changes and consistent QoS support for mobile users. Finally, we conclude this article with a discussion of further work of the QoS control for NGN.

**Index Terms**—IMS, NGN, QoS control, RACF, RACS, service

## I. INTRODUCTION

A variety of networks currently provide disparate services using different architectures and policies. These networks provide separate vertically integrated services. The next generation network (NGN) is an IP based packet switched network that provides a single network capable of carrying any and all services [1]. Unlike circuit switched networks such as the public switched telephone network (PSTN), IP networks originally lacked control mechanism for quality of service (QoS) since they were designed to provide best effort delivery without QoS [2]. QoS can be mainly guaranteed via two ways; a simple mechanism by over-provisioning in a node and a link bandwidth sharing mechanism under connection oriented architectures. The former is extremely simple since it does not require a QoS architecture. However, it does not guarantee service level agreement (SLAs) since it is possible for the buffer occupancy at an output port of a router to fluctuate widely even if the average utilization is kept low. On the other hand, the latter is complex and requires signaling, but it does guarantee QoS. Accordingly, we narrow our scope to the link bandwidth sharing mechanism under connection oriented architectures. There are commonly used schemes for QoS such as differentiated services (DiffServ) and multi-protocol label switching (MPLS) have enabled the use of QoS in IP networks, a necessary requirement for NGN.

NGN broadly consists of a service and a transport stratum, and its QoS control is executed via both strata. In the service stratum, QoS parameters are negotiated when a multimedia service is being established. NGN can employ the IP multimedia subsystem (IMS) of 3GPP as a common signaling platform for the establishment of services because of its flexibility in introducing new services using standard interfaces and

standardized service capabilities [3]. IMS consists of three planes, namely, the application plane, the IMS control plane (or core IMS), and the IP transport plane. The IMS control plane is used in the service stratum [4]. Resource control and admission based on the negotiated QoS parameters are mainly executed in the transport stratum, and the resource admission control function (RACF) [5], the resource access control system (RACS) of ETSI/ TISPAN [6], and IMS of 3GPP (i.e., IP transport plane) [7] can be employed for this purpose.

In addition, the QoS control of NGN can be further strengthened by meeting recent requirements which can be categorized into two parts. One is adaptive QoS modification according to users' QoS changes [8]. Another is QoS continuity in order to guarantee service consistency for mobile users in heterogeneous access networks [9]. Various QoS control approaches satisfying these requirements have been proposed in order to improve the QoS control of NGN.

In this article, we review and classify standards and research on the QoS control of NGN. First, we review the standardized QoS control architectures of NGN via the two strata, including RACF, RACS, and IMS. We also compare them with each other in terms of QoS coverage and number of required signaling messages. Subsequently, we give a literature review of QoS schemes for adaptive QoS modification according to users' QoS changes and QoS continuity for mobile users. Finally, we conclude this article and discuss further work for QoS control of NGN.

## II. QoS CONTROL ARCHITECTURES OF STANDARDIZATION BODIES

In order to establish a NGN multimedia service, a session setup has to take place first. This is a procedure where users negotiate their own QoS parameters with each other to establish a service, query necessary resources based on the negotiated QoS parameters, and confirm other users' resource reservation. QoS control in NGN is also executed during this session setup procedure through the service and transport strata. Below, we describe the QoS control architectures for each stratum.

### A. Service Stratum

QoS parameter negotiation is executed in the service stratum by employing several IMS functional entities (FEs) such as call session control functions (CSCFs) and a home subscriber server (HSS), as illustrated in the box entitled "Service stratum" in Fig. 1. A session among users is established by using the session initiation protocol (SIP) and the session description protocol (SDP). QoS parameters are specified in SDP in terms

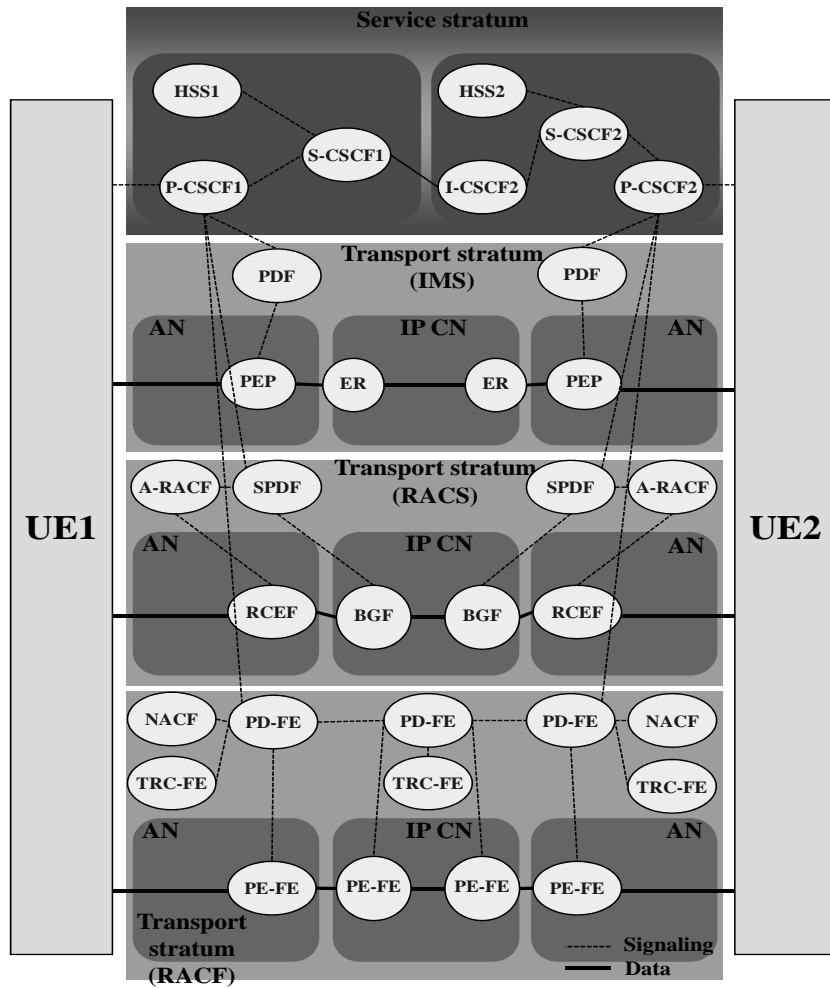


Fig. 1. The QoS control architecture of NGN with IMS, RACS, and RACF.

of required bandwidth, type of media, transport protocol, type of codes, and delay, loss requirements.

Let us consider two users as shown in Fig. 1, namely user equipment 1 (UE1) and user equipment 2 (UE2), and let us assume that they belong to different service providers (i.e., different home domains). Let us assume that UE1 wants to establish a multimedia connection to UE2. For this, UE1 has to establish first a session with UE2. Accordingly, UE1 sends a SIP message (i.e., INVITE) to the proxy-CSCF1 (P-CSCF1) in which it specifies its QoS parameters. P-CSCF1 authenticates UE1, checks the security of the SIP message, and forwards it to its S-CSCF1 which authorizes the multimedia service requested by UE1 based on the service policy and the registration status of UE1 stored in HSS1. Then, S-CSCF1 forwards the SIP message to the interrogating-CSCF2 (I-CSCF2) which is an entry point FE to the home domain of UE2. I-CSCF2 searches for the S-CSCF2 that controls UE2 and forwards the SIP message to it. S-CSCF2 in turn forwards the SIP message to UE2's P-CSCF2. UE2 receives the SIP message from its P-CSCF2. UE2 sends a response SIP message (e.g., 183 Session Progress) to UE1 via the same IMS signaling path by specifying its own desired QoS parameters. This SIP message exchange between UE1 and UE2 is repeated

until the QoS parameters are finally determined.

When P-CSCF1 forwards a response SIP message from UE2 to UE1 with the final negotiated QoS parameters, it consults with the policy decision point (PDP) in the transport stratum to verify the resource availability for the negotiated QoS parameters. If the PDP grants the necessary resources to UE1, P-CSCF1 forwards the SIP message to UE1 by informing it that it can make a resource reservation in the transport stratum. When UE1 starts its resource reservation, it sends a SIP message (i.e., PRACK) to UE2 so that UE2 can also start to make its resource reservation after sending a response SIP message (i.e., 200 OK). The remaining SIP message flow of a session establishment can be found in [4].

### B. Transport Stratum

Resource control and admission between a P-CSCF of a UE and a PDP as well as resource reservation by the UE during a session setup are executed either via IMS, or RACS, or RACF in the transport stratum. They are indicated in Fig. 1 by a separate box comprising the access network (AN) of UE1, the AN of UE2, and the IP core network (IP-CN). The appropriate FEs are also indicated.

RACF, RACS, and IMS are all policy based since policy

based QoS management provides complete network control, including traffic congestion management, traffic shaping and policing, bandwidth control and traffic balancing [10]. Resource control and admission are commonly implemented by two FEs, namely, the PDP and the policy enforcement point (PEP) [11]. A PDP is co-located with a P-CSCF. According to recent 3GPP releases, a charging rule is added to a PDP such that it is referred to as a policy and charging rules function (PCRF) [12]. A PEP is mostly located in the edge router of a UE's AN. When a PDP receives a resource request from a P-CSCF after the final QoS parameter negotiation, it determines how many resources should be allocated according to the registration status of the UE, network resource availability, and network policy. Then, the PDP maps the negotiated QoS parameters specified in SDP to specific QoS parameter semantics which are used in the transport stratum. Then, the PDP sends them to the PEP of the AN that serves the UE in order to enforce and allocate the determined resources to the UE when the UE makes a resource reservation. In summary, resources are primarily requested by a P-CSCF after QoS parameter negotiation, and they are finally allocated when they are requested by a UE. Although RACF, RACS, and IMS are all policy based, their QoS coverage, terminology, and corresponding FEs are different from each other, as summarized in Table 1.

1) *IMS*: In IMS, the PDP is known as the policy decision function (PDF), and the PEP depends on the type of AN where a UE is located. For the universal mobile telecommunication system (UMTS) in 3GPP IMS release 5, a UE can request resources via the following two FEs: the serving general packet radio service (GPRS) support node (SGSN) and the gateway GPRS support node (GGSN). GGSN acts as a PEP as well as the edge router (ER) of AN. In 3GPP IMS release 6, the AN is extended to wireless local area network (WLAN) and worldwide inter-operability for microwave access (WiMAX). The PEP in WLAN and WiMAX is known as the packet data gateway (PDG). Thus, a UE can request resources from a PDG, which consults with a PDF, via a wireless access gateway (WAG). Furthermore, in 3GPP IMS release 7, a UE is served by a digital subscriber line (DSL), and it requests resources from a broadband access server (BAS), which acts as a PEP, via a DSL access multiplexer (DSLAM). Unlike the previous two releases, release 7 has been standardized to couple with RACS, and thus a BAS is connected to the PDP of RACS.

2) *RACS*: The PDP in RACS is known as the service policy decision function (SPDF), and it authorizes resource requests of UEs received from a P-CSCF based on network policy. SPDF also consults with the access-resource and admission control function (A-RACF) which provides the functionality for admission control and resource reservation [10]. If the resources requested by a UE are available, and the UE is authorized, a SPDF informs the border gateway function (BGF), which also acts as the ingress ER of IP-CN, in order to enforce call admission and resource allocation for the UE. Simultaneously, a A-RACF informs the resource control enforcement function (RCEF) which also acts as the ER of the AN and allocates adequate resources to the UE.

3) *RACF*: The PDP in RACF is known as the policy decision function entity (PD-FE), and the PEP as the policy enforcement functional entity (PE-FE). Compared to IMS and RACS, RACF can employ an additional PD-FE for the IP-CN to control the PE-FEs which also act as ingress and egress ERs in the IP-CN. The PD-FE of each AN connects to a P-CSCF as in the case of the PDPs of IMS and RACS. However, the PD-FE covering the IP-CN is assumed to be connected to other IMS FEs located in the home domain of one of the UEs such as S-CSCF. Hence, the primary resource requester can be one of P-CSCFs in both ANs or one of the other IMS FEs in the home domain. Also, RACF introduces another FE, the transport resource control functional entity (TRC-FE) which monitors network topology changes and the amount of available resources and provides its PD-FE with the information in order to improve network resource efficiency.

According to a UE's capabilities, RACF supports two modes of resource control and admission, namely, the push and the pull modes [2]. A UE, which is capable of QoS signaling such as resource reservation protocol (RSVP), can support both modes. Otherwise, a UE can only support the push mode. Common procedures for the two modes are as follows. When the IMS FE of a home domain requests resources from the PD-FE of the IP-CN using a UE's negotiated QoS parameters, the PD-FE checks the resource availability from the TRC-FE of the IP-CN. Then, the PD-FE requests resources from the two PD-FEs of the ANs. Each AN PD-FE firstly authorizes its corresponding UE with the help of the network attachment control function (NACF). In the case of the push mode, the PD-FE of the IP-CN requests that the PE-FEs of the IP-CN allocates resources, and each AN PD-FE also informs its corresponding PE-FE to enforce resource allocation for its corresponding UE. If the necessary resources for both IP-CN and the two ANs can be guaranteed, then the resources along the path between the two UEs are reserved. In the case of the pull mode, the PD-FEs of the IP-CN and the ANs do not request their corresponding PE-FEs to reserve resources for the UE. Instead, a UE requests resources from these PE-FEs using signaling protocol such as RSVP. Each PE-FE allocates resources after it consults with its corresponding PD-FE which has been already informed of the resource request of the UE. Resources can be more efficiently utilized in the pull mode since the time duration when a PE-FE reserves resources for a UE is shorter in the pull mode than in the push mode, resources can be more efficiently utilized in the pull mode [2].

### C. Comparison of IMS, RACS, and RACF

In this section, we compare IMS, RACS, and RACF in terms of QoS coverage and the number of required signaling messages. In terms of QoS coverage, IMS provides the narrowest QoS control among the three QoS architectures. End-to-end QoS in IMS may be difficult to guarantee given that traffic congestion may occur in the ER of the IP-CN. If a UE is capable of RSVP, end-to-end QoS can be also guaranteed in IMS. In addition, it has been proposed in [11] that the PDF can control the QoS of the ER of an AN as well as that of the IP-CN as in the case of RACS. RACS can support end-to-end

TABLE I  
A COMPARISON OF IMS, RACS, AND RACF.

Type		PDP	PEP	Other FEs	Coverage/ E2E QoS	Number of signaling messages
IMS	R5 (UMTS)	PDF	GGSN	SGSN	AN/ E2E QoS in single-domain if UE capable of RSVP	Lowest
	R6 (WLAN)		PDG	WAG		
	R7 (DSL)	SPDF	BAS	DSLAM		
RACS		SPDF	RCEF (AN) / BGF (IP CN)	A-RACF	AN and IP CN border/ E2E QoS in single-domain	Medium
RACF		PD-FE	PE-FE	NACF, TRC-FE	AN and IP CN/ E2E QoS in multi-domain	Highest

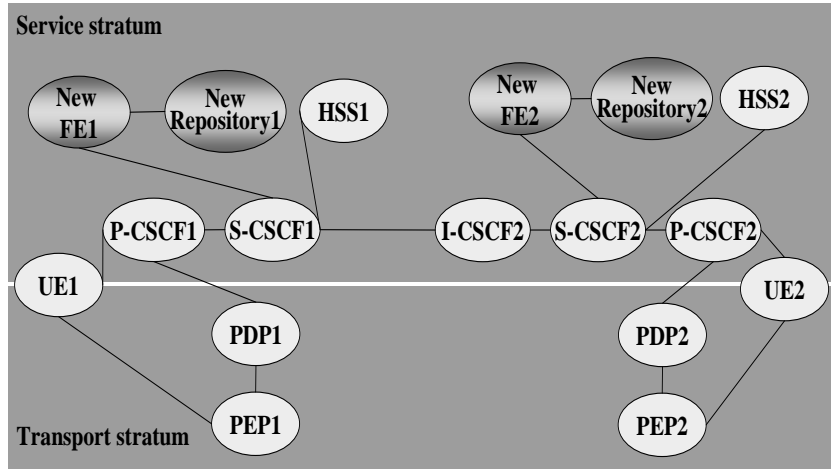


Fig. 2. The IMS NGN architecture for adaptive QoS modification.

QoS by covering the ERs of the ANs and the IP-CN. However, RACS can only control the QoS of a single IP-CN, and it cannot guarantee the end-to-end QoS in the case of multi-domain IP-CN. Compared with RACS, RACF can provide end-to-end QoS even in multi-domain of IP-CN because it can support multiple PD-FEs and PE-FEs in IP-CN, one set per domain

The required number of signaling messages has an inverse relationship with QoS coverage and depends on the signaling path. That is, the more signaling paths exist, the higher is the number of signaling messages. When a multimedia service is initiated or modified due to a UE's service change, or UE's mobility, RACF results in more signaling messages for QoS control than RACS or IMS because more FEs are involved in the resource control and admission. Furthermore, if the QoS control fails because one of the PD-FEs does not have enough resources for the negotiated QoS parameters, QoS parameter negotiation is repeated in the service stratum, and thus the signaling for resource control and admission in the transport stratum is also repeated. As a result, the volume of signaling in RACF can increase more than that in RACS and in IMS. Naturally, the time to establish a session in RACF can be greater than that in RACS or in IMS because of the higher number of signaling messages. This can be exacerbated in the case of multimedia application such as a multiparty conference

where multiple UEs are involved.

### III. ADAPTIVE QoS MODIFICATION ACCORDING TO CHANGES

In this section, we review recently proposed QoS control approaches for NGN that focus on adaptive QoS modification. UE1 can change its QoS during an ongoing session with UE2 by sending a re-INVITE SIP message to UE2 with its new QoS parameters. The subsequent signaling flow is the same as that of a session establishment. The new QoS requested by UE1 can change after the newly requested QoS parameters are negotiated and the corresponding resources are allowed by its PDP.

However, these QoS modification procedure may not make an efficient use of the network resources, and a user's needs may not be always satisfied. The following two cases identify a need for further improvement for QoS modification.

- Case I: It is better to suggest the next-best QoS to a UE than just to reject a QoS change request when the resources corresponding to the requested QoS is unavailable. In addition, even if the requested QoS is available, it may be economical to provide the UE with an optimized QoS solution that satisfies user experience and cost-effectiveness at the same time.

- Case II: A UE can adaptively determine QoS by being notified of QoS availability before it starts a session. Also, a UE regularly receives QoS availability information from its service provider such that it can efficiently change QoS using the QoS availability information during an ongoing session.

The above Cases I and II have been addressed in several papers by commonly supplementing the service stratum with a new FE and a new repository, as illustrated in the box entitled “Service stratum” in Fig. 2. The proposed solutions are classified into Type I and II corresponding to Case I and II, respectively. Type I solutions provide adaptive QoS modification during session establishment or during an ongoing session, and Type II before session establishment or during an ongoing session. Below, we review some of the proposed solutions.

#### A. Type I

In [8], the authors propose a new FE for adaptive QoS modification, referred to as the QoS parameter matching and an optimization application server (QMO AS). Let us now assume that UE1 wants to establish a multimedia connection to UE2. As described previously, UE1 sends an INVITE SIP message to UE2 with its QoS parameters. The SIP message is received by S-CSCF1 via P-CSCF1. As shown in Fig. 2, before S-CSCF1 sends the SIP message to I-CSCF2, it forwards it to the new FE1, namely, QMO AS1 (i.e., the QMO AS for UE1), which determines feasible QoS parameters based on UE1’s previous user and service profiles stored in HSS1. The SIP message containing the new QoS parameters modified by QMO AS1 is received by UE2 via the IMS signaling path S-CSCF1  $\Rightarrow$  I-CSCF2  $\Rightarrow$  S-CSCF2  $\Rightarrow$  P-CSCF2. UE2 sends a response SIP message by specifying its QoS parameters, which is received by S-CSCF2 via P-CSCF2. S-CSCF2 also forwards the SIP message to QMO AS2 (i.e., the QMO AS for UE2) in order to obtain feasible QoS parameters for UE2, and the SIP message with the new QoS parameters modified by QMO AS2 is sent to UE1. This procedure is repeated until the QoS parameters are finally negotiated between the two UEs. When the two S-CSCFs receive a response SIP message specifying the final negotiated QoS parameters, they forward the SIP message to their QMO AS which provides optimized QoS parameters using an optimization process. Both P-CSCFs query their PDP for resource availability corresponding to the optimized QoS parameters. The remaining SIP procedures are the same as in the session establishment procedure. We note that in the above example we described how QoS modification can take place during a session establishment. The above procedures can also be applied during an ongoing session.

In [13], the new FE and the new repository are referred to as the session prioritization function (SPF) and the context information base (CIB), respectively. The QoS modification is on the basis of preemption according to a UE’s service class. When UE1 establishes a session with UE2, an INVITE SIP message including its QoS parameters, is sent to S-CSCF1 via P-CSCF1. As in [8], S-CSCF1 forwards this SIP message to SPF1 (the SPF for UE1), and SPF1 checks whether UE1 can use the requested QoS parameters according to its

service class status which is stored in CIB1 (the CIB for UE1). If the requested QoS parameters are available, SPF1 informs S-CSCF1 which then sends the SIP message to I-CSCF2. The remaining procedures are the same as in the basic session establishment. If the QoS parameters requested by UE1 are not available, SPF1 investigates how to meet these QoS parameters, and specifically, SPF1 forces S-CSCF1 to send a REFER SIP message to other lower priority UEs which belong to same service provider as UE1. This REFER SIP message forces the lower priority UEs to change their QoS by sending a re-INVITE SIP message. When they finish their QoS modification, they inform SPF1 of their session re-establishment, and subsequently SPF1 instructs S-CSCF1 to proceed with the session establishment of UE1. When UE2 receives a SIP message from UE1, it sends a response SIP message by specifying its desired QoS parameters to S-CSCF2 via P-CSCF2. S-CSCF2 also forwards this SIP message to SPF2 (the SPF for UE2) which may also perform the same procedure as SPF1. The proposed QoS modification scheme can be also extended to the case of an ongoing session.

#### B. Type II

Type II, which mainly focuses on investigating QoS availability before a session is established or during an ongoing session, also uses a new FE and a repository. Compared with Type I, the new FE does not directly modify the QoS of a UE, but it enables a UE to adaptively determine its QoS parameters by providing the necessary QoS information. To do this, a UE should register with the new FE using the same procedure for registering with an application server. After this registration, the UE periodically publishes its QoS availability information to the new FE by sending a PUBLISH SIP message. The new FE uses the updated QoS availability information from all registered UEs in order to provide them with QoS availability information when they request it. Hence, a UE can request QoS availability when it wants to change its QoS during an ongoing session. Also, other UEs can obtain the QoS availability information of the UE when they want to establish a session with the UE. This can reduce the time and signaling for QoS negotiation.

In [14], a new FE referred to as the service adaptation server (SAF) and a new repository referred to as the media independent handover information server (MIIS) are introduced. When UE1 wants to establish a session with UE2, it first requests from SAF1 (the SAF for UE1) the QoS availability corresponding to its desired service. SAF1 provides feasible QoS parameters after consulting with MIIS. The QoS availability of UE2 located in another home domain can be obtained by SAF1 from SAF2 (the SAF for UE2). Since UE1 knows the QoS availability of UE2, the time to negotiate QoS parameters can decrease. UE1’s QoS change during an ongoing session can be also successful after obtaining QoS availability from SAF1. In [15], a similar approach as in [14] is proposed except that the new FE is referred to as the resource availability server (RAS).

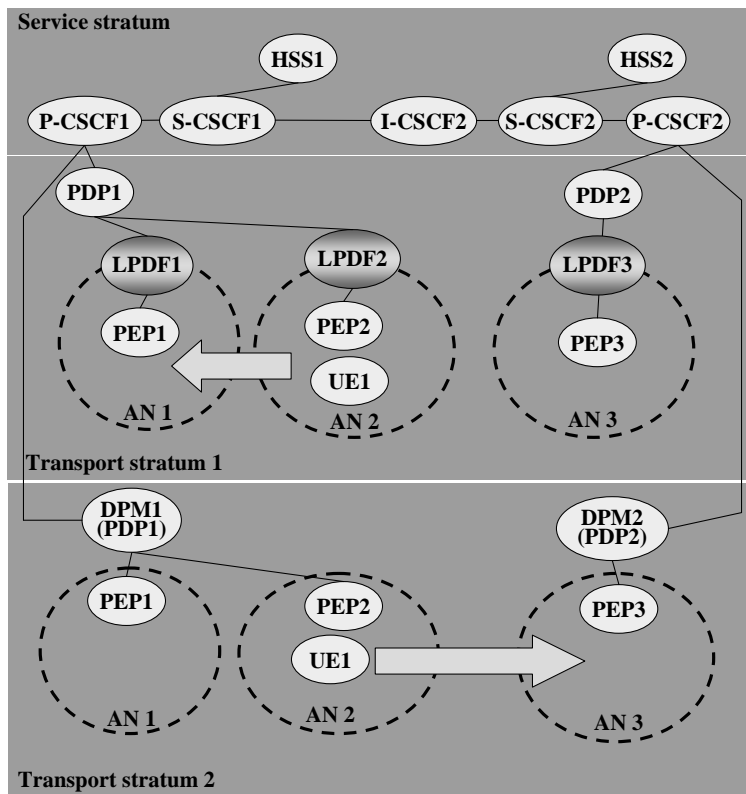


Fig. 3. The IMS NGN architecture for Type I consistent QoS support.

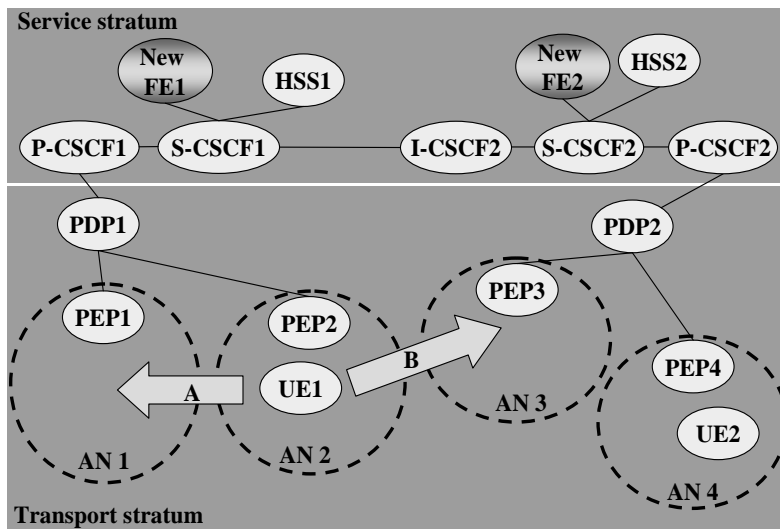


Fig. 4. The IMS NGN architecture for Type II consistent QoS support.

#### IV. CONSISTENT QoS SUPPORT FOR MOBILE USERS

There are two types of mobility, namely, terminal mobility and session mobility. The former deals with a UE's capability to be adaptive to network changes originating from mobility by supporting mobile IP and equipment modules for various network types (e.g., WLAN, WiMAX). The latter deals with service continuity without loss of QoS as a UE moves into a different AN. In this article, we survey session mobility which is supported in both strata of NGN.

In general, it is hard to guarantee consistent QoS for mobile

UEs due to the heterogeneity of ANs which may be managed by the same PDP or by different PDPs. When a UE moves into a different AN, QoS disagreement may occur due to absence of QoS arbitration and mapping. In view of this, several schemes have been reported in the literature which guarantee service continuity with consistent resource support for mobile users. These schemes can be also classified into two types, namely, Type I and II, according to the layer where QoS continuity actually takes place. The procedures to guarantee QoS continuity in Type I and II are executed in the transport

stratum and in the service stratum, respectively. As in the case of the adaptive QoS modification, new FEs have been introduced in both strata in order to maintain QoS consistency, as illustrated in Figs. 3 and 4.

#### A. Type I

In Type I, consistent QoS is supported by employing a new FE in the transport stratum. As shown in Fig. 3, the three heterogeneous ANs, namely AN1, AN2, and AN3, are handled by different PDPs. Specifically, AN1 and AN2 are managed by PDP1, and AN3 is managed by PDP2. In the case that ANs with different network technologies and policies are under the same PDP, the PDP acts as a QoS arbitrator in order to maintain the QoS of mobile UEs. To do this, a mobile UE should register with the PEP of its new AN which consults with the PDP for QoS consistency of the UE. However, as the number of ANs belonging to the same PDP increases, this arbitration due to the mobility of UEs may become a bottleneck. In [16], a new FE referred to as the local PDF (LPDF) was introduced in order to relieve the burden on the PDP, as illustrated in the box entitled “Transport stratum1” in Fig. 3. Each AN has its own LPDF which exchanges policies with the PDP and enforces the corresponding PEP of AN. For example, when UE1 enters AN1 from AN2, it sends a registration message to PEP1, which is forwarded to LPDF1 which in turn requests the QoS information of UE1 used in its previous AN, i.e., AN2 from PDP1. PDP1 sends a response message to PEP1 via LPDF1 so that the equivalent resources that UE1 used in AN2 can be allocated by PEP1.

In [17], continuous QoS support for mobile UEs in different ANs belonging to different PDPs is supported by a new FE, named the domain policy manager (DPM), which is an extended version of PDP. A DPM consists of a DPM management tool that acts as a PDP, a policy repository, a resource monitoring function to measure resource availability, and a resource negotiation function which communicates with other DPMs in order to reach QoS agreement for a mobile UE. As illustrated in the box entitled “Transport stratum 2” in Fig. 3, UE1 moves into AN3 which is under DPM2. UE1 sends a registration message to PEP3 of AN3, which is forwarded to DPM2 which in turn communicates with its previous DPM, i.e., DPM1, in order to obtain UE1’s QoS information. DPM2 sends this information to PEP3 in order to guarantee the QoS continuity of UE1.

#### B. Type II

In Type II, QoS continuity is implemented in the service stratum. To do this, a new FE is employed in the service stratum as shown in the box entitled “Service stratum” in Fig. 4. No new FEs are added in the transport stratum. A UE can visit an AN which is controlled by a new P-CSCF or the same P-CSCF. In Fig. 4, the arrow A gives the case where the P-CSCF is the same, and the arrow B where the P-CSCF is new.

The scheme proposed in [9] considers the case where a UE enters a new AN controlled by a different P-CSCF. To pursue QoS continuity for a mobile UE, the QoS parameters of the

UE are transferred from the previous P-CSCF to the new P-CSCF during an ongoing session or session establishment. For instance, let us consider the case shown in Fig. 4 where UE1 moves into AN3 controlled by P-CSCF2. Now, let us assume that UE1 wants to establish a session with UE2 which is also governed by P-CSCF2. UE1 sends an INVITE SIP message to P-CSCF2 by attaching the address of its previous P-CSCF, P-CSCF1. Before P-CSCF2 forwards the SIP message to S-CSCF2, it sends the SIP message to P-CSCF1 to obtain the previous QoS information of UE1. When P-CSCF1 receives the SIP message, it also requests the QoS information of UE1 from PDP1. PDP1 maps the QoS information in SDP and sends a response to P-CSCF1 which in turn sends a response SIP message to P-CSCF2 with the SDP attached. The remaining signaling is the same as in the session setup procedure. Thus, QoS continuity can be guaranteed by signaling between the new and the previous P-CSCF. This proposed scheme in [9] does not need any additional new FE for pursuing QoS consistency.

The scheme proposed in [18] is also SIP based and considers the case where a UE enters a new AN under the same P-CSCF. For this purpose, a new FE in the service stratum named back-to-back user agent (B2BUA) AS is introduced. When UE1 entering AN1 (arrow A), it sends a SIP message such as an INVITE or a re-INVITE to B2BUA AS via the same P-CSCF, P-CSCF1 specifying the QoS parameters that were used in the previous AN, AN2 (referred to in [18] as  $SDP_{old}$ ). Then, the B2BUA AS forwards the SIP message to the media resource function controller (MRFC) which is an IMS FE and can provide UE1 with new QoS parameters adapted to AN1 (referred to in [18] as  $SDP_{new}$ ). When the B2BUA AS receives the SIP message with  $SDP_{new}$  from the MRFC, it forwards the SIP message to S-CSCF1. The remaining signaling is the same as in the session setup procedure. This approach aims at minimizing the application-level handoff delay due to mobility and QoS adaptation.

In [19], it is also assumed that a UE enters a new AN which is managed by the same P-CSCF, and QoS consistency is supported but not during an ongoing session. Two FEs are introduced, namely the host agent (HA) in the transport stratum and the cross layer module in the service stratum. The HA acts as a PDP and the cross layer module correspond to a new FE in Fig. 4, respectively. For example, when UE1 moves into a new AN, AN1, it sends its change of location to its HA of AN1 via PEP1. Then, the HA sends the location change information (referred to in [19] as binding updates) to the cross layer module using a SIP message such as UPDATE. Then, the cross layer module informs S-CSCF1 of the binding updates. S-CSCF1 updates the location of UE1 and checks the QoS availability for the new location when UE1 starts a session.

## V. CONCLUSIONS AND DISCUSSIONS

Several standardization bodies introduced QoS control architectures for NGN, including RACF by ITU-T, RACS by ETSI/ TISPAN, and the IP transport plane of IMS by 3GPP. These QoS control standards are mainly employed in the

transport stratum of NGN for resource control and admission. In this paper, these standards were surveyed and compared in terms of QoS coverage and the number of required signaling messages. We also surveyed several recent papers which aim at strengthening present QoS control standards for adaptive QoS modification with respect to users' QoS change and QoS continuity for mobile users.

Despite exiting standards and recently published papers, the QoS control in NGN still requires further improvements. For instance, the new FEs introduced in these papers for adaptive QoS modification (e.g., QMO AS or SPF) may result in increased signaling at the expense of efficiency. By embedding the new FEs into an existing IMS FE such as the S-CSCF, we can reduce the required amount of signaling as well as the time for QoS modification. In addition, heterogeneous ANs use different sets of QoS classes and QoS parameters. A possible simplification, as described in [20], is to support universal QoS classes. In addition to these issues, we need to take into account the high complexity and scalability of the QoS control mechanism as well as the reliability and security of IP-CN, as mentioned in [2]. Accordingly, an extensive analysis of these performance metrics for QoS control mechanisms needs to be also executed as further work.

## VI. ACKNOWLEDGEMENTS

This work was supported by the Korea Research Foundation Grant funded by the Korean Government(MOEHRD) (2007-357-D00178).

## REFERENCES

- [1] A. R. Modarressi and S. Mohan, "Control and management in next-generation networks: challenges and opportunities," *IEEE Commun. Mag.*, vol. 38, No. 10, pp. 94-102, Oct. 2000.
- [2] J. Song, M. Y. Chang, S. S. Lee, and J. Joung, "Overview of ITU-T NGN QoS control," *IEEE Commun. Mag.*, vol. 45, No. 9, pp. 116-123, Sep. 2007.
- [3] M. Melnyk, A. Jukan, and C. Polychronopoulos, "A cross-layer analysis of session setup delay in IP multimedia subsystem (IMS) with EV-DO wireless transmission," *IEEE Trans. Multimedia*, vol. 9, no. 4, pp. 869-881, June 2007.
- [4] G. Camarillo and M. A. Garcia-Martin, *The 3G IP multimedia subsystem (IMS)*, Wiley, 2006.
- [5] ITU-T Rec. Y.2111, "Resource and admission control functions in NGN."
- [6] ES 282 003 V1.1.1, "Resource and admission control sub-system (RACS); functional architecture."
- [7] 3GPP TS 23. 228 V8.2.0, "Technical specification group services and system aspects; IP multimedia subsystem (IMS); Stage 2."
- [8] L. Kapov, M. Mosmondor, O. Dobrijevic, and M. Matijasevic, "Application-level QoS negotiation and signaling for advanced multimedia services in the IMS," *IEEE Commun. Mag.*, vol. 45, No. 7, pp. 108-116, July 2007.
- [9] T. Renier, K. L. Larsen, G. Castro, and H. P. Schwefel, "Mid-session macro-mobility in IMS-based networks," *IEEE Vehicular Technol. Mag.*, vol. 2, No. 1, pp. 20-27, Mar. 2007.
- [10] I. Vidal, J. Garcia, F. Valera, I. Soto, and A. Azcorra, "Integration of a QoS aware end user network within the TISPAN NGN solutions," *IEEE ECUMN*, pp. 1-9, 2007.
- [11] W. Bohm and P. Braun, "Policy based architecture for the UMTS multimedia domain," *IEEE NCA*, pp. 1-11, 2003.
- [12] 3GPP TS 23. 207 V8.0.0, "End-to-end quality of service (QoS) concept and architecture."
- [13] M. E. Barachi, R. Glitho, and R. Dssouli, "Context-aware signaling for call differentiation in IMS-based 3G networks," *IEEE ISCC*, pp. 789-796, 2007.
- [14] J. W. Floroiu, M. Corici, B. J. Lee, S. Lee, S. Arbanowski, and T. Magedanz, "A vertical handover architecture for end-to-end service optimization," *IEEE MWCS*, pp. 1-5, 2007.
- [15] T. Ozcelebi, I. Radovanovic, and M. Chaudron, "Enhancing end-to-end QoS for multimedia streaming in IMS-based networks," *IEEE ICSNC*, pp. 1-5, 2007.
- [16] M. Mani and N. Crespi, "Inter-domain QoS control mechanism in IMS based horizontally converged networks," *IEEE ICNS*, pp. 1-5, 2007.
- [17] F. C. Gouveia and T. Magedanz, "A framework to improve QoS and mobility management for multimedia applications in the IMS," *IEEE ISM*, 2005.
- [18] C. Balakrishnan and K. Al-Begain, "Towards a user-centric and quality-aware multimedia service delivery implementation on IP multimedia subsystem," *IEEE NGMAST*, 2007.
- [19] L. Le and G. Li, "Cross-layer mobility management based on mobile IP and SIP in IMS," *IEEE WiCom*, 2007.
- [20] P. Levis, M. Boucadair, P. Morand, J. Spencer, D. Griffin, G. Pavlou, and P. Trimintzios, "A new perspective for a global QoS-based Internet," *J. Commun. Software and Systems.*, vol. 1, No. 1, pp. 13-23, Sept. 2005.