A framework for improved QoS evaluation and control in SIP-based NGN

F.Weber[1,2], W.Fuhrmann[3], U.Trick[2], U.Bleimann[3], B.Ghita[1]

[1] Network Research Group, University of Plymouth, Plymouth, United Kingdom
[2] Research Group for Telecommunication Networks, University of Applied Sciences Frankfurt/M., Frankfurt/M., Germany
[3] University of Applied Sciences Darmstadt, Darmstadt, Germany
e-mail: weber@e-technik.org

Abstract

Today's standardised approaches for the control of QoS (Quality of Service) in NGN (Next Generation Networks) come along with a high volume of additional, unscalable traffic for the allocation and reservation of network resources within the IP transport network. This paper describes a new framework for comprehensive QoS control in SIP-based (Session Initiation Protocol) NGN, addressing the shortcomings of standardised NGN QoS provision and thus, leading to a more efficient QoS provision model. Because this framework approach does not rely on traditional IP QoS mechanisms it can be applied to arbitrary combinations of IP transport technologies. In order to save up network resources and, at the same time, provide appropriate service performance to the subscribers multiple factors (such as individual QoS requirements of different media and codecs) are considered to perform scalable end-to-end QoS monitoring, rating, and control.

Keywords

NGN, SIP, QoS, IP, admission control, framework for comprehensive QoS control

1. Introduction

The concept of NGN as defined by ITU-T NGN GSI (International Telecommunication Union – Telecommunication Standardization Sector NGN Global Standards Initiative) and ETSI TISPAN (European Telecommunications Standards Institute Telecoms & Internet converged Services & Protocols for Advanced Network) can be outlined by several main key features (Trick and Weber, 2007); (ITU-T Y.2001, 2004), one of which is the provision of Quality of Service (QoS).

Traditional ways to provide QoS in IP transport networks are usually based on mechanisms such as IntServ/RSVP (Integrated Services / Resource Reservation Protocol) (IETF RFC 1633, 1994) (IETF RFC 2205, 1997), DiffServ (Differentiated Services) (IETF RFC 2474, 1998) (IETF RFC 2475, 1998), or MPLS (Multi-Protocol Label Switching) (IETF RFC 3031, 2001). However, all these mechanisms come with individual characteristics that, depending on the overall design and dimension of a respective NGN architecture, potentially lead to inefficient or insufficient QoS provision, respectively. IntServ/RSVP, while supporting precise

end-to-end per-flow service provisioning, lacks of scalability (Simmonds and Nanda, 2002), (Welzl and Mühlhäuser, 2003), is considered potentially insecure (Giordano *et al.*, 2003), and, as mentioned in (Bohnert *et al.*, 2007), is not an adequate solution for the use in complex scenarios. On the other hand, the more scalable DiffServ mechanism, by definition providing only relative prioritisation for packets of selected data flows, is not efficient in dealing with network overload and is considered impractical for the use in networks that have to deal with a relative high volume of high priority traffic (Menth, 2006). Also the application of MPLS, which offers potentially beneficial properties to network operators in order to provide QoS in NGN, is limited related to scalability and efficiency (Bohnert *et al.*, 2007). Even though the combined application of at least two of the above-mentioned mechanisms generally leads to an improved relationship between scalability and efficiency, the individual issues of the mechanisms used can still have an effect (Welzl and Mühlhäuser, 2003), (Giordano *et al.*, 2003). Additionally, typical IP QoS approaches are (by definition) not aware of communication sessions (e.g., a Voice over IP session) established by higher layer protocols such as SIP. Therefore, within the standardised NGN QoS concept according to (ETSI ES 282 001, 2005) and (ITU-T Y.2001, 2004), a logical relationship between the SIP service control layer and the IP transport layer is created in order to provide QoS for the exchange of media session data between user end systems. Unfortunately, this technique comes along with a high volume of additional, unscalable traffic within the IP transport network (Park and Kang, 2005). Thus, today's standardised NGN approaches for QoS control led to inefficient resource management traffic.

2.   Standardised NGN architecture and QoS provision

The following sections describe the standardised general NGN architecture, according to ETSI TISPAN and ITU-Ts NGN GSI, and its QoS provision concept.

2.1. Standardised NGN architecture

The standardised NGN architectures defined by ETSI TISPAN in (ETSI ES 282 001, 2005) and ITU-T in (ITU-T Y.2001, 2004) correspond to each other in all fundamental aspects. Both architectures can generally be divided into two strata, the service stratum (layered on-top) and the transport stratum below.

An NGN's service stratum consists of logical functional components, allowing a subscriber to use services and applications provided by the NGN, such as (in the simplest case) initiating a SIP telephony session with another subscriber. In terms of SIP, the service stratum comprises of a SIP server infrastructure.

The NGN's transport stratum provides IP connectivity and QoS-based IP transport to the user. The IP transport functions consist of any arbitrary IP transport infrastructure, including both access and core networks.

The user equipment (e.g., a SIP end system) is connected to an interface (e.g., a DSL-based access network (Digital Subscriber Line)) of the respective NGN's transport functions. By transmitting IP packets over this interface, the user

equipment uses SIP to communicate with the NGN's service stratum (e.g., to setup media sessions to other users' end systems).

2.2. QoS provision in NGN

Once a service is requested by an NGN subscriber a top-down resource handling is performed in order to provide QoS for the respective service (such as a VoIP session). That is, if a service request originating from a user equipment enters the service stratum, resource and QoS requirements for the respective service are identified and handed down into the transport stratum where network resources are authorised, allocated, and reserved within the respective NGN's IP transport network. Depending on the IP transport technology (such as Ethernet, ATM, MPLS), the QoS mechanisms supported by the IP infrastructure (such as DiffServ or IntServ), and the network's dimension and topology, the steps required in order to provide QoS to media sessions implicate the use of a variety of different signalling protocols (such as RSVP for resource reservation), in conjunction with potentially required gateways.

3. Identified issues of NGN QoS provision and resulting requirements

As denoted in chapter 2, because of the fact that both the NGN's service stratum and the transport stratum are involved in the QoS provision process, signalling between these two strata is compulsory. Combined with the application of traditional IP QoS mechanisms (see chapter 1), this leads to an unpredictable volume of resource and QoS management traffic.

Note that the absolute amount of signalling effort in order to provide per-session QoS finally depends on the subscribers' session behaviour, because the standardised NGN QoS architecture works on a per-session basis. That is, a subscriber establishing and cancelling many sessions in a short period of time will require a substantial amount of resources and QoS management traffic.

Based on our earlier research work, the following main requirements for the provision of QoS in SIP-based NGN have been identified (Weber *et al.*, 2007).

- Functions and mechanisms, leading to a trustworthy QoS for each established session and, at the same time, do not occupy resources on a per-session basis themselves.
- Simple and resource saving QoS control should be preferred. If possible, approaches should rely on already standardised protocols (such as SIP) and architectures (such as NGN according to (ITU-T Y.2012, 2006) and (ETSI ES 282 001, 2005)).
- NGN QoS control has to be aware of a certain amount of traffic that is not session-based (such as TCP web traffic).
- The QoS provision in NGN should be independent of underlying transport technologies such as MPLS, ATM, and VLAN. Arbitrary IP network architectures should be supported, regardless their specific integrated QoS mechanisms.

4. Integrated framework for comprehensive QoS control in SIP-based NGN

Based on our research a framework has been created that fulfils the requirements stated in chapter 3 and, hence, solves the identified issues of the standardised NGN QoS. The framework combines several QoS provision requirements within telecommunication networks. As a result it leads to a comprehensive solution that can be implemented into the standardised NGN architecture according to (ETSI ES 282 001, 2005) and (ITU-T Y.2012, 2006) with no or minor changes, depending on the arrangement of the respective NGN's logical components.

4.1. Framework tasks

This subsection specifies the tasks to be fulfilled by the integrated framework for comprehensive QoS control in SIP-based NGN.

The main prerequisite to satisfy the requirements stated in section 3 is that any task fulfilled by the integrated framework for comprehensive QoS control causes only a minimum of additional traffic (such as QoS signalling traffic). Additionally the fulfilment of the tasks must lead to scalability of the QoS control framework with the number of subscribers and their behaviour, and with network characteristics (such as IP transport technology, IP QoS mechanisms, and network topology).

In order to satisfy the requirements stated in chapter 3, the integrated framework for comprehensive QoS control in SIP-based NGN is designed to fulfil the following tasks.

- Task 1: QoS measurement - Periodical evaluation (and near future prediction) of QoS conditions on every relevant segment of the transport network (e.g. between two access networks)
- Task 2: Advanced Admission Control - Integration with admission control for SIP-based services (Advanced Admission Control)
- Task 3: Manipulation of QoS conditions within the IP network (QoS control)

4.2. Framework overview and elements

Figure 1 gives an overview of the integrated framework for comprehensive QoS control in SIP-based NGN.

The transport network of the NGN, used as an example for the framework implementation shown in figure 1, consists of an IP-based NGN core network and several IP access networks. Note that the framework implementation is independent of the IP transport technology (such as Ethernet or WLAN in the access and ATM or MPLS in the core, respectively). Further on, the framework is independent of any arbitrary QoS mechanism supported by core or access networks.

For session control, in the simplest case, a SIP call server is provided by a SIP service provider. Note that SIP session control could also be performed by an IMS (IP Multimedia Subsystem) as defined by 3GPP (Third Generation Partnership Project) (3GPP TS 23.228, 2006). The SIP session control functions are usually connected to the IP core network.
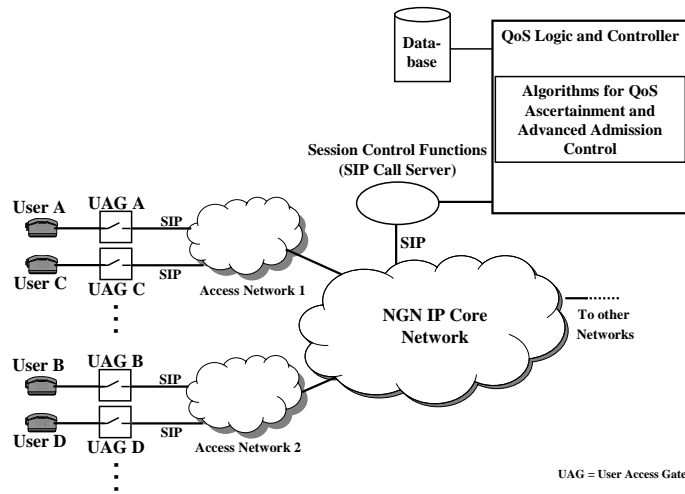
Figure 1: Integrated framework for comprehensive QoS control in SIP-based NGN (Weber and Trick, 2008)

To provide a basic NGN with the integrated framework for comprehensive QoS control, the following elements have to be added.

- User Access Gates (UAGs): This entity can be located either at the subscribers' residence or at the border of the access network. It works as a mediation device for all data exchanged between a user end system and the network and covers the functionalities shown in figure 2. Note that the UAG must be trusted by the SIP service provider.
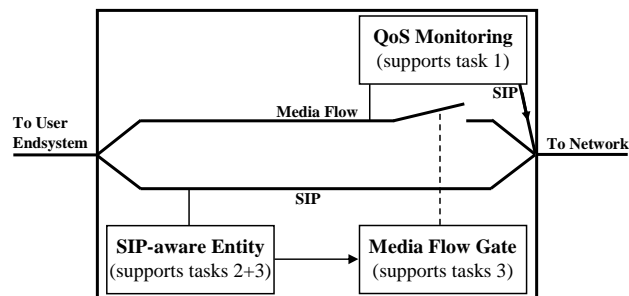
Figure 2: User Access Gate (UAG) block diagram (Weber and Trick, 2008)

- QoS Logic and Controller (QoS L&C): This centralised entity has to be provided with interfaces to the SIP session control function and to a database. It covers the functionalities shown in figure 3.
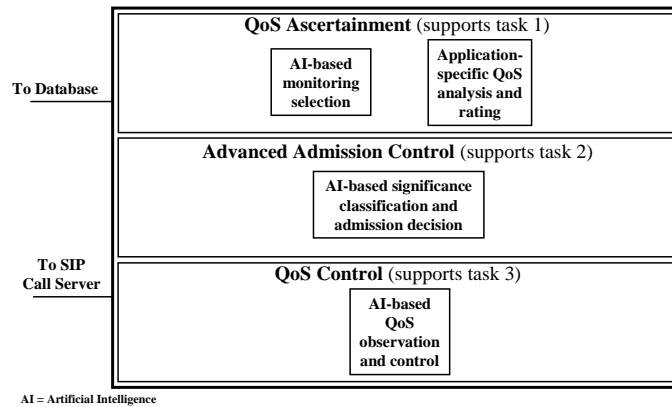
Figure 3: QoS Logic and Controller (QoS L&C) block diagram (Weber and Trick, 2008)

Within the following chapters, the UAG and QoS L&C entities are described in the context of the overall framework.

4.3. Framework basic principle

As denoted in chapter 2, the standardised NGN QoS architecture makes use of QoS and resource control within and below the IP layer, applying a top-down QoS control scheme. This leads to the inefficiencies and the lack of scalability mentioned in chapter 3.

The integrated framework for comprehensive QoS control does not deal with active QoS control on or below the IP layer. It is assumed that the arbitrary IP core and access networks within the NGN may provide best effort IP transport. Therefore, two subscribers A and C, with the same connectivity and access network (see figure 1), will experience the same QoS when connecting to two other subscribers, B and D, from a remote network, providing that both sessions use the same parameters (i.e. both sessions are based on the same medium and use the same codecs). In this case, in order to collect information about the QoS experienced by subscribers A and C, it is sufficient to obtain QoS information for one of these two sessions.

4.4. Framework functionality fulfilling task 1

In the simplest case, the information provided by the QoS Monitoring functionality included in the UAG (see figure 2) comprises jitter and packet loss rates for the IP packets carrying the media data to the user end system. This information can be obtained by the QoS L&C, prompting the SIP Call Server to query the UAG for the QoS information of an ongoing media session.

To save IP transport capacity the QoS Ascertainment functionality included within the QoS L&C (see figure 3) keeps track of all ongoing sessions in order to select carefully the UAGs that have to be queried for QoS information (note that comparable sessions under comparable conditions between comparable subscribers will experience the same QoS). The QoS Ascertainment functionality makes use of artificial intelligence (AI) in order to learn which UAGs are queried best in order to obtain a sufficient overview of the QoS between several access networks, considering several different media and codecs, and to minimise the number of queried UAGs.

The relative QoS information retrieved from each queried UAG is matched with the QoS L&C through QoS matrixes, providing information about QoS values (such as jitter and packet loss rates) tolerated by the respective medium and codec. This matching leads to the information about the absolute QoS experienced by the respective subscriber. The QoS matrixes are stored within a database that can be accessed by the QoS L&C.

By continuously keeping record of the QoS experienced for several sessions of different media between different subscribers, the QoS Ascertainment functionality of the QoS L&C creates timeline-based QoS profiles for every network segment (e.g. between two different access networks). The timeline-based QoS profiles can be used for near-future QoS prediction by other sub-functionalities of the QoS L&C.

4.5. Framework functionality fulfilling task 2

When a session is requested with or by a subscriber of a NGN, a SIP request is sent to the SIP session control functions. In order to consider the given QoS situation for the admission control, the SIP Call Server provides all relevant information about the requested session (such as the involved subscribers' identities, type of media and codecs, affected network segments) to the Advanced Admission Control functionality of the QoS L&C (see figure 3). The Advanced Admission Control functionality performs the following steps in order to decide how the session request has to be handled.

- Step 1: Media session significance classification. The objective significance of each media session is classified considering the following parameters, specific to the involved subscribers.
  - General subscriber policy criteria (e.g. premium versus basic customer)
  - Formerly experienced session availability ratio (policy-assured/granted) per medium per subscriber

  If applicable, further parameters may be taken into account for significance classification. A ranking list of significance for all ongoing and requested sessions is continuously kept and updated.
- Step 2: Identification of QoS conditions required for media session. By matching the requested media and codecs with QoS matrixes stored within the database, the Advanced Admission Control identifies QoS parameter values (such as jitter and packet loss rate) that will be tolerated by the respective media and codec.

- Step 3: Identification of network segments of the path between the endpoints.
- Step 4: Analysis of current and prospective near-future QoS conditions within the affected network segments. The Advanced Admission Control functionality obtains this information from the QoS Ascertainment functionality.
- Step 5: Admission Decision. Taking into account all relevant information derived from the performance of the steps 1 to 4, the Advanced Admission Control functionality makes a decision about the requested media sessions. The following results are possible.
  - The session will be rejected
  - The session will be granted as-is
  - The session will be granted under QoS downgrade conditions (e.g. a low-bitrate codec has to be used)

  In order to consider the ranking list of significance (see step 1), in case that a concerning incoming session request potentially would not experience satisfying QoS conditions if granted (due to high traffic volume in the network) the Advanced Admission Control can also decide to reject or downgrade competing media sessions that are objectively less relevant. For the achievement of the goal to cancel or reject as less (objectively least crucial) sessions as possible in order to maintain or recover sufficient QoS conditions for as much (objectively more crucial) sessions as possible, Artificial Intelligence (AI) will be investigated as a potential alternative for mathematical calculation as a basis for the Advanced Admission Control functionality of the QoS L&C.

In any case the admission decision is provided to the SIP Call Server in order to react adequately on the SIP session request and, if applicable, cancel or downgrade concurrent sessions. On its way to user end systems, the SIP signalling originated from the SIP Call Server is identified and recognised by the SIP-aware Entity within the UAGs (see figure 2) of the respective subscribers. The UAGs take responsibility that the subscribers' user end systems follow the signalled directives (e.g. codec downgrade or session rejection/cancellation). The SIP-aware Entity controls the UAGs' Media Flow Gate functionalities (see figure 2) (e.g. for bandwidth limitation or media flow cut-off) in order to enforce the QoS L&C's decision.

4.6. Framework functionality fulfilling task 3

As stated in chapter 4.4 the QoS Ascertainment functionality of the QoS L&C collects, derives and arranges particular information of the QoS experienced by subscribers connected to several access networks, exchanging several kinds of media coded with several codecs. This information can be accessed by the QoS Control functionality of the QoS L&C (see figure 3) in order to continuously observe the overall QoS provided for several network segments. Also the timeline-based QoS profiles are considered by the QoS Control functionality in order to predict near-future trends of the overall QoS.

In case a potential or yet existing shortage of QoS is detected for one or several ongoing media sessions (e.g. in a high traffic period), the QoS Control functionality

has to react in order to maintain or recover satisfying QoS conditions for as much high priority media sessions as possible. Hence, the QoS Control functionality can decide to cancel or downgrade concurrent media sessions that are objectively less relevant. Therefore, the ranking list of media session significance (see chapter 4.5, step 1) is taken into account. The action that has to be performed to make decisions, and the involvement of SIP call server and UAGs in order to execute and control codec downgrades and SIP session cancellations are similar as described within step 5 of chapter 4.5, and the paragraph below step 5, respectively.

5. Conclusions and Future Work

This paper proposed an integrated framework for comprehensive QoS control in SIP-based NGN. In contrast to the standardised NGN QoS architecture, this framework does not rely on top-down network resource allocation, but makes use of algorithms and artificial intelligence (AI) in order to provide sufficient QoS conditions for as much objectively crucial sessions as possible. This is achieved without any impact on the IP transport network or on any QoS mechanisms provided by the IP layer or the underlying IP transport technology.

The next step in order to bring the framework to implementation state will be the definition of the basis (such as neural networks) for the required artificial intelligence, and define the respective algorithms. Further, a prototype of the framework will be planned and implemented into an existing NGN under laboratory conditions. In future the framework may be extended to the active manipulation of the volume of non-session-based traffic (such as TCP web traffic) carried over the network in order to provide sufficient QoS to media sessions. For this the SIP-based control of the UAGs may be a key functionality.

6. Annotation

7. References

3GPP TS 23.228 (2006), Technical Specification, "IP Multimedia Subsystem (IMS); Stage 2 (Release 5)", 3GPP

Thomas Michael Bohnert, Edmundo Monteiro, Marilia Curado, Alexandre Fonte, Michal Ries, Dmitri Moltchanov, and Yevgeni Koucheryavy (2007), "Internet Quality of Service: A Bigger Picture", 1st OpenNet Workshop - Service Quality and IP Network Business: Filling the Gap, Diegem/Brussels, Belgium, March 2007

ETSI ES 282 001 V1.1.1 (2005), ETSI Standard, "NGN Functional Architecture Release 1", ETSI TISPAN

ETSI TS 185 001 V1.1.1 (2005), Technical Specification, "NGN Quality of Service (QoS) Framework and Requirements", ETSI TISPAN

Silvia Giordano, Stefano Salsano, Steven Van den Berghe, Giorgio Ventre, and Dimitrios Giannakopoulos (2003), „Advanced QoS Provisioning in IP Networks: The European Premium IP Projects", *Communications Magazine*, IEEE, Volume 41, Issue 1, January 2003, pp. 30 - 36

IETF RFC 1633 (1994), Request For Comments, "Integrated Services in the Internet Architecture: an Overview", IETF

IETF RFC 2205 (1997), Request For Comments, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", IETF

IETF RFC 2474 (1998), Request For Comments, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", IETF

IETF RFC 2475 (1998), Request For Comments, "An Architecture for Differentiated Services", IETF

IETF RFC 3031 (2001), Request For Comments, "Multiprotocol Label Switching Architecture", IETF

ITU-T Y.2001 (2004), Recommendation, "General overview of NGN", ITU-T

ITU-T Y.2012 (2006), Recommendation, "Functional requirements and architecture of the NGN release 1", ITU-T

Michael Menth (2006), „Das echtzeitfähige und ausfallsichere Internet der nächsten Generation – Next Generation Networks" (The real-time capable and fail-proof Internet of the next generation – Next Generation Networks), lecture series on ‚The Internet and its applications', BTU Cottbus, Germany, 25 April 2006

Juyoung Park and Shin Gak Kang (2005), "QoS Architecture for NGN", *Advanced Communication Technology, ICACT 2005*, pp. 1064-1067, IEEE, July 2005

Andrew Simmonds and Priyadarsi Nanda (2002), „Resource Management in Differentiated Services Networks", IFIP Interworking 2002, *Proceedings in 'Converged Networking: Data and Real-time Communications over IP'*, pp. 313 - 323, ed. C McDonald, pub. Kluwer Academic Publishers, ISBN: 1-4020-7379-8, 2003

U. Trick and F. Weber, (2007), *SIP, TCP/IP und Telekommunikationsnetze*, Oldenbourg, Munich, Germany, ISBN: 978-3-486-58228-4

F. Weber, W. Fuhrmann, U. Trick, U. Bleimann, B. Ghita, (2007), "QoS in SIP-based NGN – state of the art and new requirements", *Proceedings of the third collaborative research symposium on Security, E-learning, Internet and Networking SEIN 2007*, pp. 201-214, Information Security & Network Research Group – University of Plymouth, Plymouth, ISBN: 978-1-8410-2173-7

F. Weber and U. Trick, (2008), "Optimizing and simplifying SIP-based NGNs' QoS architecture", International SIP 2008 Conference, Paris, France

Michael Welzl and Max Mühlhäuser (2003), „Scalability and Quality of Service: A Trade-off?", *Communications Magazine*, IEEE, Volume 41, Issue 6, June 2003, pp. 32 - 36