

# Enhancing end-to-end QoS for multimedia streaming in IMS-based networks

**Citation for published version (APA):**

Ozcebebi, T., Radovanovic, I., & Chaudron, M. R. V. (2007). Enhancing end-to-end QoS for multimedia streaming in IMS-based networks. In *Proceedings 2nd International Conference on Systems and Networks Communications (ICSNC 2007) 25-31 August 2007, Cap Esterel, France* (pp. 48-48). IEEE Computer Society. <https://doi.org/10.1109/ICSNC.2007.35>

**DOI:**

[10.1109/ICSNC.2007.35](https://doi.org/10.1109/ICSNC.2007.35)

**Document status and date:**

Published: 01/01/2007

**Document Version:**

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

# Enhancing End-to-End QoS for Multimedia Streaming in IMS-Based Networks

Tanır Özçelebi, Igor Radovanović, Michel Chaudron

Eindhoven University of Technology, Dept. of Mathematics and Computer Science  
System Architecture and Networking, P.O. Box 513, 5600 MB, Eindhoven, The Netherlands.  
(t.ozcelebi, i.radovanovic, m.r.v.chaudron)@tue.nl \*

## Abstract

*Convergence of the emerging IP Multimedia Subsystem (IMS) includes unlicensed, nondedicated and nondeterministic, hence uncontrollable computer access networks for IP multimedia services. It enables provision of resource demanding real-time services and multimedia communication raising new end-to-end Quality-of-Service (QoS) challenges, for which quality adaptation using resource management is proposed as a solution in this paper. This is an integrated solution taking both IMS and computer access networks into account as well as the two end-devices and the application servers (AS) involved in communication. The best user experience is targeted under real-time variation of available network (e.g. bandwidth, buffer space) and end-device (e.g. battery, CPU, memory, storage) resources throughout a session. The multimedia content is dynamically adapted to fit the resource availability variations, achieving maximum system (i.e. network and end-devices) resource utilization and enhanced QoS. The resource availability update signalling is carried over Session Initiation Protocol (SIP) during the session. This is work in progress.*

## 1 Introduction

The rapid growth of the mobile telecommunication society since the first launch of Global System for Mobile Communications (GSM) [1] had been continuously sustained until recently. However, this growth is slower than ever nowadays, especially in the developed countries due to widespread use of broadband Internet access. A simple personal digital assistant (PDA) with network connectivity is now enough to access voice, video and data services outdoors over the Internet. Unlike the traditional mobile telecom services, these services do not always come with guaranteed quality and a best-effort approach is usually tolerated by the Internet users. Such services still continue to attract

interest from the mobile telecom community and upgrading from plain voice services to data and multimedia services was seen necessary. Among these services, real-time multimedia streaming and multimedia messaging can be listed as the most promising ones, which are currently enabled by the emerging 3<sup>rd</sup> Generation (3G) mobile technologies such as High-Speed Downlink Packet Access (HSDPA) [2], Evolution-Data Optimized (1xEV-DO) [3] and Universal Mobile Telecommunications System (UMTS) [4].

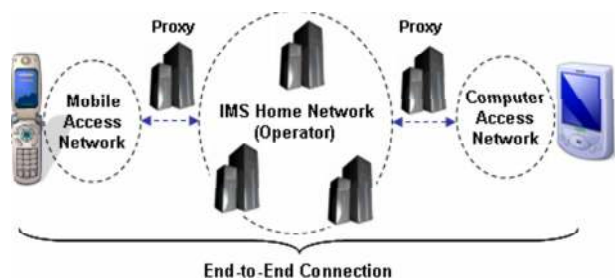
During the evolution of telecom standards from GSM towards 3G up until now, telecom operators have been the owners of the wireless frequency band licenses and the network infrastructure for their services, including access points, i.e. base stations (BS). Thus, QoS issues can be resolved much easier within a completely licensed mobile domain (e.g. GSM) by monitoring the available network resources for each communication session. Here, a communication session can be defined as a durable connection between two parties over the session layer. Furthermore, guaranteed QoS policies such as admission control (AC) [5], resource reservation [6] and traffic engineering [7] can be enforced for the offered services, so that the system does not have to serve more users than its capacity at a given time. One problem here is that all service types offered for certain end-device types are always assumed to be available disregarding the end-device resources. Moreover, mobile communications is currently going towards an All-IP network direction, which is accessible via not only conventional mobile telecom networks but also any IP access point including the Internet. For this reason, the converged Next Generation Networking (NGN) framework [8] was introduced recently, where it is proposed that the functionality of services be independent from the underlying network and, managed and unmanaged IP networks can be combined together.

Note that the QoS enforcement mechanisms of mobile standards (e.g. GSM, UMTS etc.) during a single session are no longer valid for NGN as the access network ownership has been taken away from the operator. For instance, applying admission control and resource reservation on NGN sessions in a non-NGN access network (e.g.

\*This work is being supported by COMET consortium funded by the European Commission 6<sup>th</sup> Framework Programme (FP6).

computer networks) would not prevent failure at peak hours due to capacity overload and interference from non-NGN flows. Therefore, another serious mobile communication challenge is raised considering heavily resource consuming multimedia streaming services, namely *end-to-end QoS*.

The IMS [9] is a backward compatible NGN architecture standardized by the 3rd Generation Partnership Project (3GPP) [10] for enabling a large variety of Internet-like services with easy provisioning for telecom operators. The Session Initiation Protocol (SIP) is employed at the application-layer as a control safeguard in order to simplify integrating IMS with the Internet. According to the IMS specifications, the users should be able to access services anytime, from anyplace and any SIP-enabled device [11]. Therefore, it is envisioned that the operators must provide services to users even in access networks that are beyond their control. The end-to-end QoS issue can be divided into two stages, i.e. *i)* QoS in nondedicated best-effort networks with dynamic device and network resource availability and, *ii)* integrated QoS over the entire end-to-end path combining both dedicated and nondedicated networks. Generally, the term “end-to-end” refers to the connection from the user-equipment (UE) or the proxy to the server in the telecom network. The path from one end to the other lies within the core network of the operator, which is owned, i.e. controlled by the operator according to this definition. On the other hand, here the term “end-to-end” is used to refer to the connection from one UE to the other, passing through the access networks and the core network as shown in Fig. 1.



**Figure 1. Proposed end-to-end connection**

Guaranteeing network resource availability is not feasible even if all parts of the communication path are operator-owned, since the access networks are mostly nondeterministic, e.g. WLAN and Ethernet. Thus, the access networks constitute uncontrollable and variable-resource bottlenecks of the end-to-end path. Furthermore, the provisioning of advanced services puts higher demands on the network and the end-device resources and the resource availability in the end-devices may become the bottleneck. As a result, despite resource reservation and admission control in the core network (CN), it will no more be possible to *guarantee* end-to-end resource availability and QoS as

mentioned in the 3GPP specifications [9]. Since guaranteed resource availability and QoS cannot be achieved at the link layer, it is necessary to enhance QoS through optimization at the other layers of the communication stack. Such techniques exist in the literature including media aware transmission such as the partial order connections scheme [12] and cross-layer optimization schemes [13]. However, none of these techniques are designed for IMS networks.

In this work, we propose a framework for enhancing the end-to-end QoS for multimedia delivery in IMS-based networks, in which user experience is enhanced with application layer adaptation. A Service Quality Management (SQM) framework that requires monitoring of available system resources is introduced, hence the need for Resource Managers (RM) and resource availability signalling. The current IMS architecture lacks in providing specifics of such QoS negotiation and adaptation on-the-fly *during the session*. In this paper, we propose such a solution without deforming the existing IMS architecture.

The rest of this paper is organized as follows: The existing QoS mechanisms in IMS and related work in the literature are explained in Section 2. Our proposed solution with resource management and QoS adaptation is described in Section 3. Finally, the conclusions are drawn in Section 4.

## 2 Existing QoS Mechanisms in IMS

The IMS standard was developed considering high-end services such as multimedia communication that may require vast amount of resources from the UE and the network with real-time constraints. Naturally, this creates the service quality issues mentioned in Section 1, which boils down to a best-effort service. However, the telecom customers are used to getting guaranteed QoS on their mobile services in general, and such customer profile cannot be convinced otherwise in an easy way.

The IMS architecture shown in Fig. 2 is divided into three planes with different functionalities [9]. The transport and connectivity layer is separated from the higher application and service layer by means of the signalling and control layer, whose job is to carry out call session control. In this plane, there is a standard set of controls valid for all services provisioned. Ideally, if there is a policy agreement among the networks that lie along the end-to-end path, the serving operator(s) can use these control mechanisms in order to improve the end-to-end QoS of the provisioned services. Due to such agreement, the operator would be aware of the available resources in these networks and the networks would apply the QoS decisions of the operator regarding call admission and resource reservation. However, even if such an agreement is present, a QoS policy translation problem across different networks still exists in practice.

The operator can employ its own policies for QoS negoti-

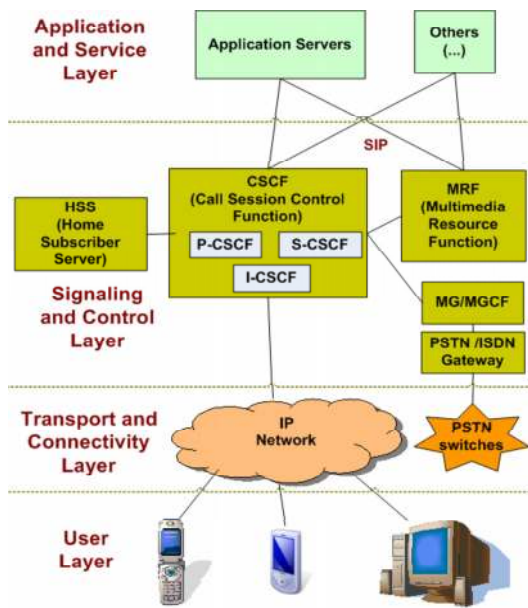


Figure 2. IMS layered architecture

ation and resource reservation within the CN at the transport layer due to ownership. However, the access networks outside the CN may have their own QoS models and semantics, which cannot be interpreted in other networks along the session path [14]. For example, there exist four different QoS classes for UMTS (i.e. conversational, streaming, interactive, background) [4], three classes for DiffServ [15] (i.e. default, expedited forwarding and assured forwarding), and three classes for IntServ [16] (i.e. guaranteed, controlled load, best effort). Moreover, many computer access networks have very limited QoS support, e.g. Ethernet, IEEE 802.11b. This non-standardization throughout different network domains makes it more difficult to establish the desired end-to-end QoS reservations and to come up with fair charging agreements on communication paths whose networks are owned by multiple parties. Maniatis et al. [14] have tried to tackle the QoS model translation problem across different networks by applying an intelligent mapping algorithm during end-to-end QoS negotiation such that the best suiting QoS class is selected in each network along the session path. Similarly, the operator can employ AC by using the Policy Decision Function (PDF) at the access network border and making the access network physical bearer enforce a Service Based Local Policy (SBLP), if the bearer actually listens to the commands from the IMS signaling and control layer.

This inter-network policy agreement assumption is somewhat superficial for access networks with unlicensed frequency band and anonymous usage rights. If there is no such agreement, the AC cannot prevent non-IMS users

from connecting to the access network and consuming its resources and already existing IMS sessions in the access network can be jeopardized. Furthermore, even such agreement would not solve the problem if the access network is nondeterministic and nondedicated, which is the case for Ethernet and WiFi. The wireless environment introduces much interference from the outer world, magnifying the nondeterministic behavior in WiFi.

In the IMS standard [9], it is specified that QoS parameters can be negotiated between two UE's prior to the session establishment. Once the QoS parameters have been negotiated between UE's (checked against fixed device capabilities) and been approved/modified by the Call Session Control Functions (CSCF) (checked against user subscription credentials) associated with both users, the IMS network asks the CN and the access network to reserve resources for this session. The SIP INVITE and SIP UPDATE messages [17] are used for this purpose as shown in Fig. 3.

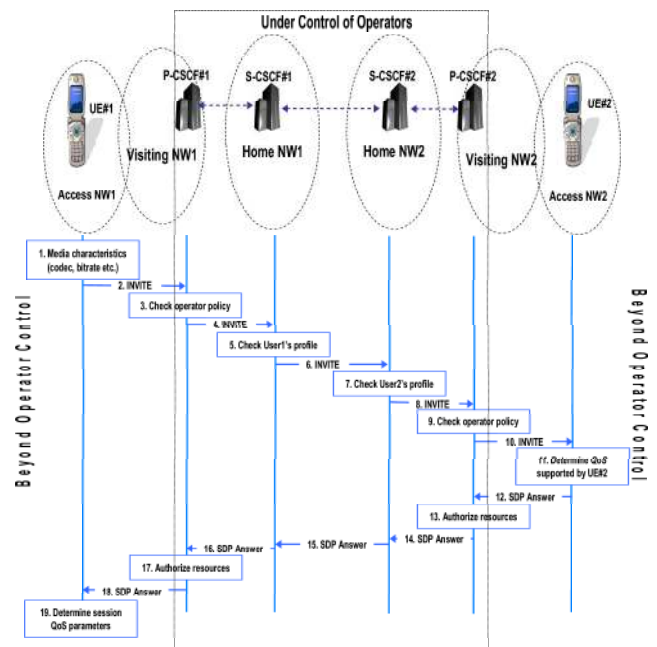


Figure 3. Pre-session QoS negotiation in IMS

The first INVITE message from the caller UE (UE1) to the call receiver (UE2) carries the QoS proposal (request) of the sender and this proposal is checked against the subscription levels of users at the Serving-Call Session Control Functions (S-CSCF) in the home networks of both UE1 and UE2, and QoS parameters are modified at these locations if there is a mismatch. Afterwards, UE2 puts her own QoS proposal in the answer and this proposal is again checked and modified at the associated S-CSCF's of both users according to their subscription status. Finally, UE1

can accept this counter QoS proposal and start the session or try to renegotiate with a SIP UPDATE message. Note that, the second negotiation is again carried out prior to the session start. Within the body of these SIP messages the session data is passed using Session Description Protocol (SDP) [18].

As mentioned previously, the IMS sessions are bound to suffer unless the session QoS parameters are continuously modified to fit the resource availability when there is a shortage or a boost. If this is not the case, overprovisioning can be employed for less resource demanding services such as the legacy SMS messaging, whose size is only 160 bytes with no real-time requirements. However, for more advanced services with extra resource requirements (e.g. video streaming), overprovisioning is infeasible as there may be many users trying to access similar content at a given instant. Note that the real-time multimedia services of today call for significant amounts of resources that are time varying to function properly. High data throughput and low signalling delay is critical at the network side, while high processing power, memory, storage space and energy is required at the client side until the end of the multimedia session. The pre-session QoS negotiation shown in Fig. 3 is not sufficient for this purpose, since *i*) the network and client resources could change drastically and unpredictably afterwards, *ii*) the session could be indefinitely long (real-time communication), and *iii*) there could be software running other than the multimedia communicator draining resources from the UE's. Session renegotiation depending on requests from the application, network load and link quality is also mentioned the IMS standard [9], while the implementation specifics of such a *QoS renegotiation* mechanism are not provided.

### 3 QoS Adaptation with Resource Management

The IMS allows multimedia sessions for which session QoS guarantees cannot be given [9]. For example, if the user is connected through a nondedicated access network, it is not possible to guarantee end-to-end QoS, deteriorating user experience. This problem can be solved by introducing service quality adaptation (SQA) according to system resource availability (local, remote and network). The design constraint for the solution is to avoid possible alterations to the IMS architecture.

In order to perform such adaptation, the proposed SQM module responsible for managing services must be aware of the availability of local (e.g. battery, memory, storage CPU etc.) and network (e.g. throughput, delay etc.) resources in the session time and adapt service quality level/type accordingly, hence the need for *resource availability monitoring* and *signalling*. In this work, we propose *i*) Resource Man-

agement (RM) and Service Quality Management (SQM) modules, *ii*) a Resource Availability Server (RAS) as an Application Server (AS), and *iii*) a resource availability signalling mechanism for real-time adaptation of multimedia communication streaming and data streaming (e.g. video-on-demand) services in the IMS network.

The RM module is a crucial part of the proposed framework. It is responsible for tracking local and network resources available to the UE in real-time. At the receiving UE, this information is published by the RM module to the RAS server to be accessed by the remote transmitting UE. At the transmitting UE, the RM module is responsible for gathering such resource availability data of the remote receiving UE from RAS. The local and remote UE and network resource availability data gathered by RM is an input to the SQM module inhabiting the transmitting UE (local) for evaluation as shown in Fig. 4.

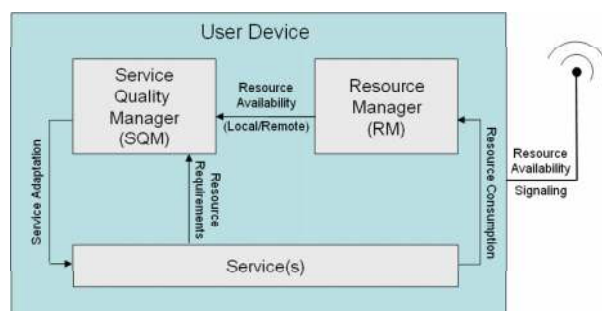
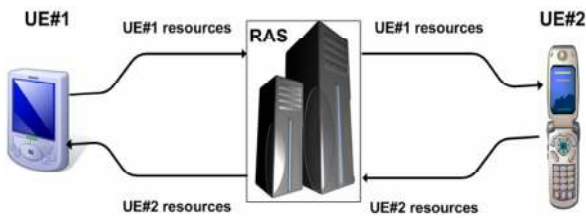


Figure 4. RM and SQM modules

The SQM module compares the required resources for different QoS levels of the session and the available resources at the network, local and remote end-devices, determines the maximum achievable quality in real time and adapts QoS accordingly. If SQM determines that even the lowest allowable quality for the multimedia session cannot be reached, it can switch to a different service type, e.g. it can switch to audio-only instead of audio plus video. It is our implementation choice that the SQM module resides in the end-user devices for load balancing purposes. Alternatively, SQM could have been built as a server inside the CN to be enjoyed by all subscribed users instead of this load-balanced and distributed approach. However, this would cause extreme computational overload on SQM server at peak hours, which is why it is not preferred in our solution.

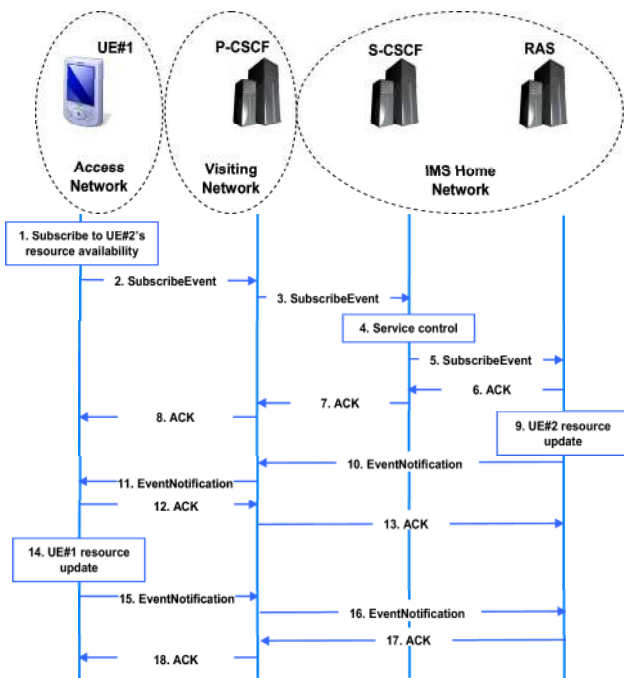
It is the proposed RAS server that is responsible for collecting resource availability information from the receiving end-devices and delivering it to the transmitting end-device. Note that an end-device can be transmitting multimedia, receiving multimedia, or both in a multimedia communication scenario, e.g. videoconferencing. The signalling of resource availability data from and to the RAS server needs to be IMS compliant. hence it is appropriate to employ the SIP

protocol for resource availability signalling in the proposed scheme as shown in Fig. 5. Here it is assumed that user equipments UE1 and UE2 had registered to each other's resource availability information at RAS.



**Figure 5. In-session resource signalling**

Resource availability data is carried from end-devices to RAS and back in SIP event notification messages, such that a resource update is signalled whenever the local resources (e.g. memory, CPU, storage etc.) or network resources (e.g. bandwidth, jitter etc.) at one end crosses a critical boundary threshold. The proposed message flow diagram from the end-device to RAS for resource availability signalling is depicted in Fig. 6 and our additions to the SIP/SDP parameters as resource indicators are shown in Table 1. After the addition of the proposed resource availability parameters, an example update message looks as shown in Table 2.



**Figure 6. Proposed in-session SIP signalling**

In this way, both users will be aware of each other's local and network resources and SQM can use this information to perform its functionality at the transmitting device

**Table 1. Proposed resource data in SDP**

r= (resource name) memory, CPU, storage, throughput, battery  
t=\* (attribute type) Mbytes, Kbytes, Percentage, etc  
a=\* (zero or more resource attribute lines)

**Table 2. Example SDP with resource data**

```
NOTIFY sip:abc.somename.com SIP/2.0
Via: SIP/2.0/UDP abcd.tue.nl:5060;branch=z9hG4bKnashds7
Max-Forwards: 70
To: Bob <sip:server.somename.com >
```

```
...
v= RM/1.45
o= abc 5876768686868 7698798797979 IP 1.2.3.4
s= 123456789
i= Resource update to presence server
c= IN IP4 1.2.3.4
b= 100 kbps
k= none
```

```
r= memory
i= free memory status
t= kbytes
a= 12450
```

```
r= battery
i= battery charge remaining
t= percentage
a= 86
```

for enhancing user experience at the receiving side. Alternatively, the SQM in the end user device may update the list of available communication services only (with respective maximum quality levels) by invoking resource to service mapping technique. The preliminary achievements of this work include implementation of the RAS module in the middle and RM module software at the end-devices (i.e. PDA's with WinCE operating system) with limited capability. We are able to establish Real-time Transport Protocol (RTP) video and audio sessions between PDA's who are subscribed to each other's resource information at the RAS server and to carry out real-time resource availability signalling using SIP/SDP in parallel with the session. An example snapshot of the RAS interface is shown in Fig. 7 with real-time resource data of both end-devices on display, i.e. remaining battery percentage (BT), storage space (ST) and dynamic memory (MM).

Our experiments with video streaming in the ad-hoc wireless IP network show that, the quality of the multimedia content at the receiving PDA changes according to distance and occlusions in between devices and interference from the environment. Sometimes the quality gets below



Figure 7. RAS log interface

the acceptable levels, such that massive visual artifacts and pauses during real-time playback are observed as expected. We are continuing our research with adaptation of multimedia on real session time, through which these quality deficiencies can be kept to a minimum for users' convenience.

## 4 Conclusion

The IMS allows users to benefit from resource consuming multimedia services at anytime and anyplace. As a result, in contrary to the traditional practice in telecom networks, the IMS access networks can be nondedicated and/or nondeterministic, and the availability of the network and the end-device resources may show severe oscillations within a session. This is a serious problem especially for streaming multimedia communication services as their resource demands and delay requirements are more stringent than that of deferred message delivery. The available QoS mechanisms in the IMS standard are insufficient in this respect, and rigorous quality degradation and annoying pauses during video playback are inevitable. In this work, resource availability signalling and QoS adaptation is proposed as a solution to this problem. Currently, the available resources, hence the accessible services, of a user are successfully advertised to the users that are subscribed to her resource status at the RAS server through SIP signalling, which is carried out in parallel with the multimedia session. The final outcome of this work will be a complete end-to-end SIP resource signaling and in-session adaptive QoS control scheme for IMS based networks.

## References

- [1] ETSI EN 301 243 v4.0.1, Dec. 1997.
- [2] "Technical Specification Group Radio Access Network; High Speed Downlink Packet Access; Overall UTRAN Description; Release 5," 3GPP, Tech. Rep. 3G TR 25.855 version 5.0.0, Sep. 2001.
- [3] IS-856-2 cdma2000 High Rate Packet Data Air Interface Specification, TIA Std., Rev. 2, Oct. 2002.
- [4] S. Chia, "The Universal Mobile Telecommunication System," *IEEE Comm. Mag.*, vol. 30, no. 12, pp. 54-62, Dec. 1992.
- [5] R. Yavatkar, D. Pendarakis, R. Guerin, "A Framework for Policy-based Admission Control," RFC 2753, IETF, Jan. 2000.
- [6] R. Braden, L. Zhang, S. Berson, S. Herzog, S. Jamin, "Resource ReSerVation Protocol (RSVP)," RFC 2205, IETF, Sep. 1997.
- [7] D. Awduche, A. Chiu, A. Elwalid, I. Widjaja, X. Xiao, "Overview and Principles of Internet Traffic Engineering," RFC 3272, IETF, May 2002.
- [8] <http://www.itu.int/ITU-T/ngn/index.phtml>, May 2007.
- [9] IP Multimedia Subsystem (IMS), Stage 2, V7.6.0, TS 23.228, 3GPP, Dec. 2006.
- [10] <http://www.3gpp.org>, May 2007.
- [11] <http://www.comet-consortium.org>, May 2007.
- [12] L. R.-Cardenas, E. Chaput, L. Dairaine, P. Senac, M. Diaz, "Transport of Video over Partial Order Connections", *Comp. Networks: The Int. Jour. of Comp. and Tel. Netw.*, vol.31, no.7, pp.709-725, Apr. 1999.
- [13] T. Ozcelebi, O. Sunay, R. Civanlar, M. Tekalp, "Cross-Layer Optimized Adaptation and Scheduling for Multiple-User Wireless Video Streaming," *IEEE J. on Sel. Ar. in Com.*, vol.25, pp.760-769, May 2007.
- [14] S.I. Maniatis, E.G. Nikolouzou, I.S. Venieris, "End-to-End QoS Specification Issues in the Converged All-IP Wired and Wireless Environment," *IEEE Comm. Mag.*, vol.42, no.6, pp.80-86, Jun. 2004.
- [15] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, "An Architecture for Differentiated Services," RFC 2475, IETF, Dec. 1998.
- [16] R. Braden, D. Clark, S. Shenker, "Integrated Services in the Internet Architecture: an Overview," RFC 1633, IETF, Jun. 1994.
- [17] J. Rosenberg, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler "SIP: Session Initiation Protocol," RFC 3261, IETF, Jun. 2002.
- [18] M. Handley, V. Jacobson, "SDP: Session Description Protocol," RFC 2327, IETF, Apr. 1998.