Analysis of Performance Issues in an IP-based UMTS Radio Access Network

Xavier Pérez-Costa
NEC Network Laboratories
Heidelberg, Germany
perez@netlab.nec.de

Kjell Heinze
NEC Network Laboratories
Heidelberg, Germany
heinze@netlab.nec.de

Albert Banchs
Universidad Carlos III
Madrid, Spain
banchs@it.uc3m.es

Sebastià Sallent
Universitat Politècnica de Catalunya
Barcelona, Spain
sallent@mat.upc.es

ABSTRACT

The substitution of ATM transport by IP in future UMTS Radio Access Networks (UTRAN) introduces several performance challenges that need to be addressed to guarantee the feasibility of its deployment. The significant increase of the overhead requires of header compression and multiplexing methods to achieve a usage of the UTRAN resources similar to the ATM one. Additionally, the specific UTRAN transport needs require the adaptation of standard packet scheduling mechanisms to efficiently use the network resources while providing the required QoS. Our results show that, applying header compression plus multiplexing techniques and taking into account the specific UTRAN synchronization requirements for QoS scheduling, very significant performance improvements can be obtained.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design-Wireless Communications; C.4 [**Computer Systems Organization**]: Performance of Systems

General Terms

Design, Performance

Keywords

IP, RAN, Iub interface, QoS, WFQ, EDF

1. INTRODUCTION

At the end of 1999, 3GPP (Third Generation Partnership Project) started working toward an All-IP architecture. The architecture evolution was driven by two main objectives: independence of the transport and control layer to ease the implementation of new applications up to the mobile terminal and operation and maintenance optimization for the access network. During the transition from 3G to All-IP systems, the UMTS Terrestrial Radio Access Network (UTRAN) and Core Network (CN) should evolve toward the IP transport solution as an alternative to the original asynchronous transfer mode (ATM) transport. The substitution of the ATM transport in the UTRAN by IP (IP-RAN) has already been proposed and specified by 3GPP for Release 5 [1], Figure 1 depicts the IP-RAN architecture. However, this substitution raises several issues in order to make an efficient use of the network that have not been fully addressed yet by 3GPP:

The protocol overhead increases significantly when IP transport is considered as compared to ATM. Mechanisms to improve the IP transport efficiency have to be engineered to guarantee the feasibility of UMTS IP-based radio access networks.

Due to the 'best effort' nature of IP, the actual level of QoS the network is able to guarantee to applications is uncertain. Methods to guarantee QoS differentiation while keeping the efficient use of the UTRAN resources are required to meet applications' needs.

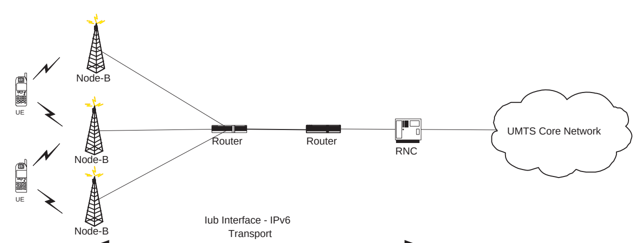

Figure 1: IP-RAN architecure

Even though the problem of properly supporting IP in the RAN is a recent topic, it has already received some attention in the literature. 3GPP standardization has started discussions on the convenience of adopting IP-RAN as can be observed for instance in the technical report [1]. A study comparing the efficiency of ATM and IP transport was performed by the MWIF (Mobile Wireless Inter-

MSWiM'05, October 10–13, 2005, Montreal, Quebec, Canada.

net Forum) and can be found in [2]. Venken et al. [3] identified the need for admission control for voice traffic in IP-RAN. Finally, the related work closer to our study that the authors are aware of is the one performed by Menth et al. [4] where different QoS scheduling mechanisms were evaluated for AAL2.

In the following sections we present our proposed solutions to alleviate the aforementioned unefficiencies. We describe first a multiplexing scheme that improves the efficiency of the transport between the RNC and the Node-Bs. Then, in Section 3, we propose a scheduling mechanism adapted to the UMTS radio access network transport requirements and application needs. Finally, a summary of the main results concludes the paper.

2. IP-RAN TRANSPORT OVERHEAD

Nowadays, the predominant traffic in current cellular networks is voice traffic. The evolution of UMTS networks toward IP transport introduces the possibility of migrating current circuit-switched voice calls to the packet-switched domain using Voice over IP (VoIP). However, in an IP-based UTRAN, the overhead required to transmit each single VoIP packet largely exceeds the actual information carried per packet resulting in an unefficient use of the scarce networks resources specially in the wireless medium and in the last-mile links.

To reduce part of this overhead header compression schemes can be considered. Robust Header Compression (RoHC), which has been specially designed for the case of wireless links, can compress the RTP, UDP and IPv6 headers together. In [5] a performance study of the header reduction achieved by RoHC was done reporting an average size for the compressed header of 6 bytes.

The RoHC solution solves the first part of the problem corresponding to the reduction of the overhead in the air interface. Unfortunately, since we can assume neither direct links from the Node-Bs to the RNC nor RoHC-capable IP routers in the path, we can not apply the same mechanism to reduce the overhead introduced by the IP transport in the UTRAN. This problem has already been studied in the literature resulting in similar solutions [2, 6, 7, 8]. The proposed solutions reduce the IP transport overhead by multiplexing several packets from different connections into a single packet. A multiplexed packet, in contrast to a packet with a compressed header, is an ordinary IP packet and can be transparently routed through the Internet.

A multiplexing scheme based in PPP multiplexing is presented in [7]. The advantage of this scheme as compared to other ones is that it is based in a standardized solution [8][1].

3. QOS FOR IP-BASED UTRAN

In this section the specific requirements of the communication between the RNC and the Node-Bs are explained followed by a description and evaluation of our proposed QoS scheduling method which aims at taking into account both the UTRAN synchronization requirements and the QoS application needs.

3.1 Iub Interface Specific Requirements

The communication between the RNC and the Node-Bs (Iub interface) has specific synchronization requirements for packets sent through a dedicated channel (DCH) as described in [9]. We focus on the downlink direction (RNC to UEs) since it is the most restrictive one. In the following, the two synchronization requirements that apply to the downlink communication are described.

[1]Due to space restrictions we had to remove these results from the paper

First, the fact that the RNC already creates the radio frames that are going to be sent by the Node-B over the air interface, i.e., assigns to each particular radio frame a connection frame number, results in a synchronization requirement which specifies that packets have to arrive at the Node-Bs at a certain instant with a maximum deviation corresponding to a specific *receiving window* value [9]. Packets arriving at an instant of time above or below this maximum deviation trigger a timing adjustment process that is used to keep the synchronization of the DCH data stream in the downlink direction (see Figure 2). If packets arrive at an instant of time above a fixed maximum deviation they are discarded since the reserved resources are not available anymore.

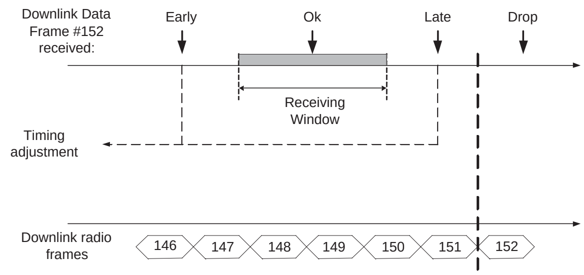

Figure 2: RNC to Node-B Synchronization

Second, the soft- and softer-handover mechanisms require the arrival of the radio frames at the UE received via different Node-Bs/sectors within a certain time interval to allow for a selection or combination of the received frames. If this synchronization requirement is not fulfilled, the quality of service experienced for this connection will be degraded and the resources used in the wired and wireless part wasted. Figure 3 depicts the soft/softer-handover mechanism.

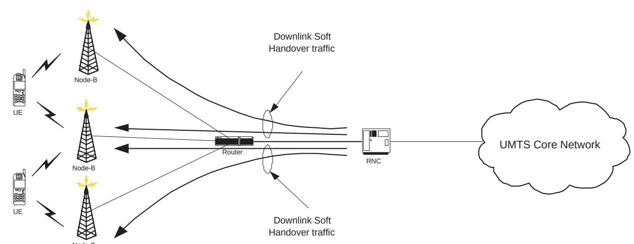

Figure 3: Soft handover synchronization requirements

In practice, these two synchronization requirements result in *all* packets sent by the RNC to the Node-Bs having a common delay deadline to meet at the same time both requirements. Therefore, *all* packets traversing the Iub interface in the downlink direction have to fulfill delay requirements.

3.2 Differentiated Packet Scheduling Design for UTRAN

Standard scheduling solutions at intermediate routers in the path from the RNC to the Node-Bs as Priority Scheduling, Weighted Fair Queuing or Earliest-Deadline-First suffer from the same major drawback: they do not take advantage of the relaxed delay requirements of low priority packets since, because of performing the differentiated scheduling once the Iub interface synchronization requirements apply, all packets have to fulfill exactly the same delay deadline and thus, it is no longer possible to delay some of them without risking their arrival on time at the Node-B. Therefore, these solutions do not perform well in our scenario since, under congestion, packets with low priority will very likely reach their destination with a delay larger than their deadline and will therefore be dropped, resulting in useless consumed bandwidth both in the bottleneck links and in the air interface.

In order to achieve the objectives stated at the beginning of this section of designing a scheduling mechanism that takes into account both the Iub interface synchronization requirements and the application QoS needs to guarantee an efficient use of the UTRAN scarce resources we design the differentiated downlink packet scheduling described in the following.

Our proposed alternative to overcome these problems is to schedule packets at the RNC according to their priorities as the receiving window constraint only starts counting once the connection frame number has been assigned by the RNC. By delaying packets at the RNC instead of at the intermediate routers a differentiation based on the different QoS application requirements is possible without the cost of risking that due to this additional delay the packets will not fulfill the Iub synchronization deadline.

We consider EDF as the best suited method to provide the desired QoS differentiation in this case, while fulfilling the Iub synchronization requirements, since an explicit reference to the time deadlines to be fulfilled is possible. Note however that similar results could be obtained with WFQ with the single drawback of requiring a higher configuration effort due to the non-explicit relationship of the assigned weights to the resulting delay.

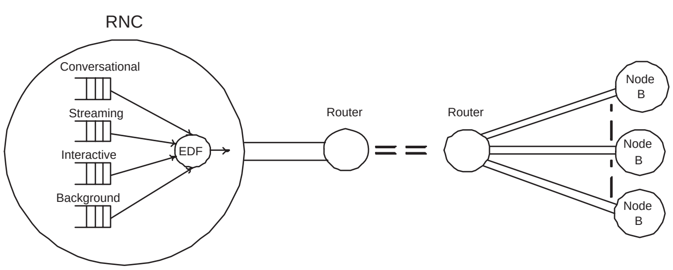

Figure 4: Differentiated Downlink Scheduling at the RNC

Figure 4 depicts our scheduling proposal. Packets arriving to the RNC to be delivered to a Node-B are classified depending on their QoS class and buffered in their corresponding queue. A deadline is assigned to each packet based on their arrival time and the traffic category. The RNC schedules then the transmission of the packets to the Node-Bs based on the deadlines of the first packet of each queue using the EDF method. If a multiplexing scheme is used, e.g., the one described in Section 2, the scheduling is applied to the multiplexed packets taking into account the deadline of the first packet inserted in the container of each traffic class. Note that the proposal relies in a proper admission control performed at the core network to guarantee that enough resources are available to fulfill the accepted services.

3.3 Performance Evaluation & Discussion

In this section we evaluate via simulation the performance improvement that results from implementing our proposed EDF downlink packet scheduling at the RNC instead of standard QoS scheduling mechanism as WFQ or simple FIFO scheduling at the bottleneck router. Based on the improvement in the performance observed in our overhead study results we consider for this study the case of using the described multiplexing scheme in the Iub interface and RoHC in the air interface. The results have been obtained with the OPNET simulator in the scenario described below. The simulations required to modify the UMTS R'99 elements provided by OPNET in order to emulate the IP-RAN architecture described in Section 1, e.g., substitution of ATM transport by IP, to implement our proposed scheduling mechanism at the RNC and the multiplexing scheme. The scenario chosen for the analysis is similar to the one depicted in figure 1. It consists of a variable number of UEs uniformly distributed along the coverage area of 5 NodeBs and of an RNC connected to the Node-Bs through a bottleneck link. The bottleneck link is emulated as one E1 link (2048 kbps capacity).

Since voice is currently the predominant application in wireless networks and is supposed to be the main application generating traffic in 3G networks in the near future, we consider two different scenarios where the mix of VoIP traffic with other multimedia traffic increases progressively. The first scenario represents a 'Mid-term' scenario where multimedia users represent 20% of the total number of users and the other 80% are VoIP ones. In the second scenario, denoted as 'Long-term', the number of VoIP users and of other multimedia traffic are equal, i.e., 50% each group.

UMTS defines four different types of classes depending on their QoS requirements [10].Therefore, we divided the multimedia users generating traffic different than VoIP as shown in Table 1. Background traffic was selected as the most significant after Conversational (VoIP) because of the wide spread e-mail usage.

Traffic Class	Conv	Stream	Inter	Backg
Mid-term	80%	5%	5%	10%
Long-term	50%	10%	10%	30%

Table 1: Distribution of the number of users of each QoS class used on each scenario.

In the following we describe the characteristics of the traffic generated by each different class:

Conversational AMR [11] is the chosen voice codec for VoIP for future IP-based UMTS networks [6]. We modeled the AMR VoIP traffic generated by a UE as a Markov chain where a VoIP source during a call session can be in the ON or OFF state, generating a rate of 12.2 kbps or 1.8 kbps, respectively. The permanence time in the ON and OFF state is 3 seconds in average.

Streaming MPEG-4 streaming of the movie Jurassic Park (average rate 35kbps, frames generated every 40ms [12]).

Interactive Web traffic, page interarrival time exponentially distributed with mean 60 seconds and an average page size of 7250 bytes.

Background Short FTP downloads, inter-request time exponentially distributed with a mean of 10s, file size exponentially distributed with a mean of 10 kbytes.

The Transmission Time Interval (TTI) considered for this study is of 20ms and the maximum delay allowed after the TTI before discarding the packet is 5ms. The configuration for the WFQ scheduler at the intermediate router has been obtained empirically through simulation, when the bottleneck link is not congested, with the single objective of providing a lower delay to the classes with a higher priority. The EDF scheduler has been configured with deadlines multiple of the TTI and different for each QoS class. The deadlines chosen based on the considerations provided in [13] and [10] are: 20ms for Conversational, 40ms for Streaming, 100ms for Interactive and 200ms for Background.

3.3.1 *Iub Interface Delay*

We investigate first the differences in the delay experienced by each QoS class when the three different scheduling mechanisms are

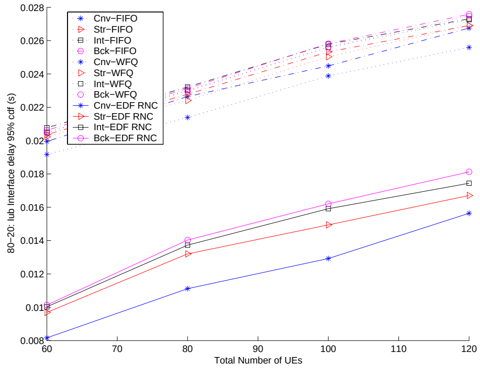

Figure 5: Impact of the number of UEs over the Iub Interface delay in the 80%-20% user mix case

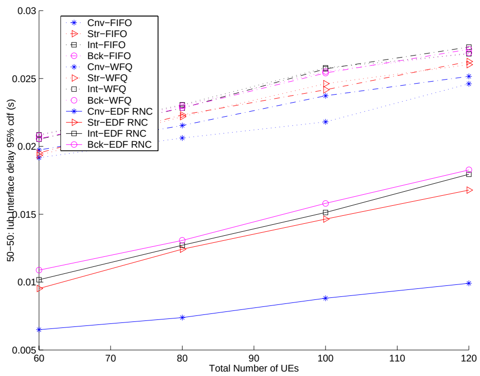

Figure 6: Impact of the number of UEs over the Iub Interface delay in the 50%-50% user mix case

used in the 80-20 user mix case. In Figure 5 we can observe that, as expected, while the FIFO scheduling provides a very similar delay for all classes the WFQ scheduling introduces differentiation between them based on the weights assigned. A small differentiation is observed though in the FIFO case due to the different frame generation rate of the applications that results in the QoS classes generating traffic less often observing a larger queue length than the ones generating more often. Even though by introducing WFQ scheduling in the intermediate router the QoS performance improves, the improvement is minor as compared to the case of performing EDF scheduling at the RNC. As commented before, the fact of performing the scheduling at the RNC instead of after leaving it allows to delay low priority packets when the Iub interface receiving window requirement does not apply yet. Additionally, since with our proposal the connection frame number is set once the packet is ready to be sent by the RNC, the Iub interface delay is lower because the multiplexing delay is excluded.

Similar results are obtained in the 50-50 user mix case, as shown in Figure 6, where the main difference now is that, due to the larger amount of lower priority traffic, a higher differentiation can be obtained by the WFQ and EDF scheduling schemes. A particular case can be observed in congestion conditions (100 and 120 UEs) for the FIFO scheduler where the the Interactive class experiences a slightly higher delay than the Background one. This is just due to our specific 50-50 user mix case (30% of Background users and 10% of Interactive ones) and application traffic characteristics

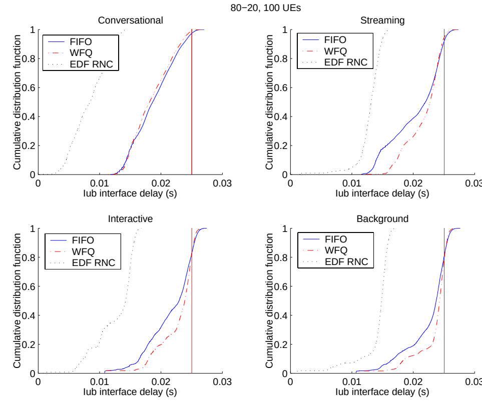

Figure 7: Impact of the number of UEs over the Iub Interface delay cdf in the 80%-20% user mix case

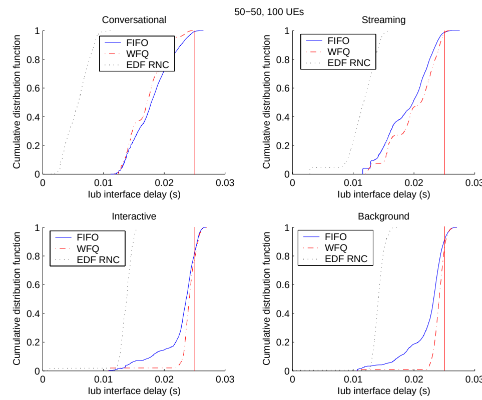

Figure 8: Impact of the number of UEs over the Iub Interface delay cdf in the 50%-50% user mix case

which results in the Background traffic observing a queue length shorter than the Interactive traffic.

In the rest of the experiments we focus on the 100 UEs case as the most relevant one for our study since it is the first case where for both user mix cases some traffic classes experience a delay above the configured maximum of 25ms before the packets are discarded.

3.3.2 Iub Interface Packet Error Rate

Figures 7 and 8 provide details on the change in the delay distribution, cumulative distribution function (cdf), for the different QoS classes introduced by the two QoS scheduling mechanisms as compared to plain FIFO. As we can observe, in the 80-20 user mix case, WFQ manages to significantly improve the FIFO performance by trading-off some additional delay for the different classes. On the other hand, in the 50-50 user mix case the WFQ configuration chosen favours in excess the conversational class resulting in a poor performance with regard to the additional delay for the rest of the classes as compared to the FIFO case. Finally, the EDF scheduling at the RNC shows a clear improvement in the delay observed for both cases with respect to the other two scheduling methods with the additional advantage of improving the control of the jitter.

The results with respect to the packet error rate observed at the Iub interface, corresponding to the delay results previously presented, are shown in Figures 9 and 10. As expected, the results perfectly match the delay distribution ones where while the EDF

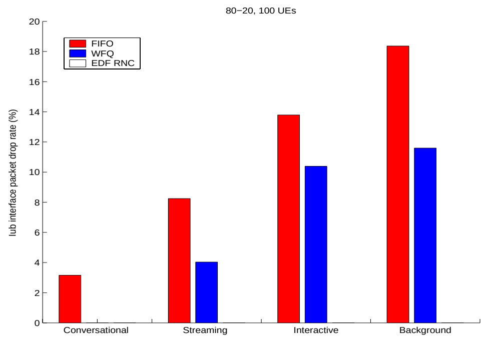

Figure 9: Impact of the number of UEs over the packet drop rate in the 80%-20% user mix case

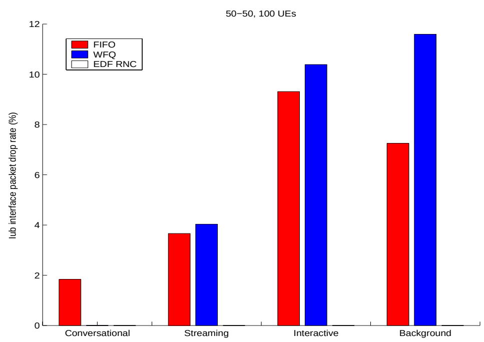

Figure 10: Impact of the number of UEs over the packet drop rate in the 50%-50% user mix case

scheduler manages to keep all the values below the dropping threshold at the Node-Bs, the WFQ scheduler improves mainly the packet losses for the Conversational class and depending on the configuration also for the rest of the classes. The particular case already observed in the delay graph for the 50-50 user mix case is reproduced here where due to the configured traffic mix for this case the losses for the interactive class are higher than for the background one.

4. CONCLUSIONS

In this paper several issues that arise when IP transport is introduced in the UTRAN have been addressed. The significant increase of the overhead requires of header compression and multiplexing methods to achieve at least a usage of the UTRAN resources similar to the one when ATM transport is used. Otherwise, the relevant increase in the network running costs due to capacity problems would threaten the actual deployment in the future of IP-based RANs. Additionally, the specific UTRAN transport requirements force the adaptation of the standard packet scheduling mechanisms to make an efficient use of the network resources while providing the required QoS.

In Section 2 header compression and multiplexing schemes for the IP transport in the Iub interface have been described which would significantly alleviate the overhead problem based on our perfromance results that we have not included due to space restictions

In Section 3 we proposed a differentiated scheduling mechanism based on EDF that fulfills the Iub interface specific requirements while providing the QoS differentiation required by the applications. The method, whose main novelty is to perform the scheduling at the RNC instead of at intermediate routers, allows to make use of the different QoS application needs without interfering with the UTRAN transport requirements resulting in a relevant improvement with respect to the Iub interface delay and the packet error rate. The main drawback of our proposal is the increase in the complexity at the RNC because of setting the connection frame number to the radio frames at the instant when the regular IP packets or multiplexed ones have to be sent.

Based on our results we can conclude that by using standard mechanisms adapted to the IP-based UTRAN specific requirements a significant enhancement of the network performance can be achieved. This enhancement though comes at the cost of a non-negligible increase of the network functionality complexity that has to be carefully considered when substituing the ATM transport by IP.

5. ACKNOWLEDGMENTS

The authors would like to thank Juan Noguera from the 3G Technologies group and Masanori Taketsugu from the Mobile Network Development Division in Japan for fruitful discussions regarding this work.

6. REFERENCES

[1] 3GPP TR 25.933, “IP Transport in UTRAN Technical Specification Group Radio Access Network,” version 5.4.0, December 2003.

[2] “IP in the RAN as a Transport Option in 3rd Generation Mobile Systems,” Mobile, Wireless Internet Forum, MTR-006, v2.0.0, 2001.

[3] K. Venken, D. De Vleeschauwer, and J. De Vriendt, “Dimensioning the Backbone Network for an IP-based UTRAN,” Proceedings of the 2001 International Conference on Third Generation Wireless and Beyond, May 2001.

[4] M.Menth, M.Schmid, H.Heiss, and T.Reim, “MEDF - A Simple Scheduling Algorithm for Two Real-Time Transport Service Classes with Application in the UTRAN,” in Proceedings of INFOCOM, March 2003.

[5] “Acticom,” http://www.acticom.de.

[6] “3GPP, IP Transport in the UTRAN Work Task Technical Report,” 3GPP TR 25.933 V5.3.0, 2003.

[7] K.Venken, I.Gomez Vinagre, and J.De Vriendt, “Analysis of the Evolution to an IP-based UMTS Terrestrial Radio Access Network,” IEEE Wireless Communications Magazine, October 2003.

[8] C. Fox R. Pazhyannur, I. Ali, “PPP Multiplexing,” IETF RFC 3153, August 2001.

[9] “Technical Specification Group Radio Access Network; UTRAN Iu interface user plane protocols,” 3GPP TS 25.415 V6.1.0, September 2004.

[10] “3GPP, Group Services and System Aspects; Services and Service Capabilities,” 3GPP TS 22.105 V5.2.0, June 2002.

[11] 3GPP TR 26.236, “Packet switched conversational multimedia applications; Transport protocols,” version 5.3.0, June 2003.

[12] F.H.P. Fitzek and M.Reisslein, “MPEG-4 and H.263 Video Traces for Network Performance Evaluation,” IEEE Network,Vol. 15,No. 6,pages 40-54, November/December 2001.

[13] 3GPP TR 25.853, “Delay Budget within the Access Stratum,” version 4.0.0, March 2001.