Content-Aware Scheduling for Video Streaming over HSDPA Networks

Luca Superiori, Martin Wrulich, Philipp Svoboda and Markus Rupp
Institute of Telecommunication and RF Engineering,
Vienna University of Technology,
Vienna, Austria
Email: {name.surname@nt.tuwien.ac.at}

Joachim Fabini
Institute of Broadband Communications
Vienna University of Technology,
Vienna, Austria
Email: joachim.fabini@tuwien.ac.at

Wolfgang Karner, Martin Steinbauer
mobilkom austria AG
Email: w.karner@mobilkom.at,
m.steinbauer@mobilkomaustriagroup.com

***Abstract*—In this article we propose the implementation of a content-aware cross-layer optimization for video streaming in HSDPA networks. The video packets are marked at the IP layer based on the impact on the quality of experience at the application layer. In the core network these packets are mapped onto multiple logical paths with different quality of service accordingly to their marking. The proposed approach complies with the 3GPP releases newer than R99. It has been tested with an HSDPA system level simulator and H-264 video traces. The results of simulations show that the proposed content aware scheduling delivers over 1 dB Y-PSNR improvement compared to the standard round robin.**

I. Introduction

Increasing available bandwidth and decreasing market prices made the internet traffic in wireless cellular networks grow significantly in the last years. The traffic consists of a variety of services and applications [1] with different requirement in terms of bandwidth, delay and jitter. This calls for a seamless convergence of the two worlds: on the one side the third Generation Partnership Project (3GPP) and on the other one the Internet Engineering Task Force (IETF).

The latest evolutions of the Universal Mobile Telecommunication System (UMTS) [2] are already oriented to an all-IP system. In the fifth release of the standard the IP Multimedia Subsystem (IMS) [8] was specified. In spite of the effort to define an all IP wireless network, packets generated by different applications still have to be handled differently by the network elements. A VoIP real time conversation between two humans has strong delay constraints: once exceeded a given time threshold the packets are of no use for the end user application. The same constraints are not in force for *background traffic*: applications like File Transfer Protocol (FTP) or file sharing are, under reasonable limits, delivery time insensitive.

Although the problem of guaranteeing Quality of Service (QoS) in UMTS has been widely discussed, it is well known that the final user is interested in the Quality of Experience (QoE). Rather then using objective parameters, such as delay or bandwidth, the customers satisfaction has to be measured observing whether the service provided by the vendor matches the expectation of the user. For specific applications, such as video streaming, particular packets have a higher impact on the end user perceived quality. In order to improve the QoE, we address the problem of differentiating the traffic within a single application data flow.

In this paper we investigate the feasibility of such a proposal in the High Speed Downlink Packet Access (HSDPA) [3] network. The video streaming is considered as a study case, but the proposed approach has not to be considered limited to this specific application. Our approach relies on the 3GPP approach of splitting the connection between the application server and the end user into multiple logical paths. An appropriate QoS class is then assigned to each logical connection, and each packet is routed via a path, depending on its relative importance for the end user QoE. By means of transmission simulation performed using an HSDPA system level simulator [4], [5], the performance of the proposed method is compared with the standard network implementation. The results are presented in terms of luminance Peak Signal to Noise Ratio (Y-PSNR), a distortion metric correlated with the quality of the video as perceived by the users.

This paper is structured as follows: Section II offers an overview of the HSDPA network. Section III introduces necessary video coding and transmission concepts. In Section IV a brief review of *state-of-the-art* QoS in UMTS is offered. The proposed method and the simulation results, are discussed in Section V and VI, respectively. The conclusions drawn in Sec. VII wrap up the paper.

II. Overview of HSPA Networks

HSDPA and its uplink pendant High Speed Uplink Packet Access (HSUPA) together are called *High Speed Packet Access* (HSPA). They form the evolution of WCDMA in wireless mobile networks. Both enhancements have lead to increased spectral efficiency in down- and uplink, with peak data rates of 14.4 Mbit/s and 4.05 Mbit/s, respectively [3].

In the context of this work we restrict ourselves to HSDPA, focusing on the cross-layer optimization of streaming in the downlink. HSDPA features a number of enhancements compared to UMTS, with the most significant ones being [6]

- a single Downlink Shared CHannel (HS-DSCH),

- fast link adaption utilizing Adaptive Modulation and Coding (AMC),
- Hybrid-ARQ retransmission handling with Incremental Redundancy (IR),
- and new MAC-hs layer including a TTI-based scheduler.

The link adaptation provides channel quality information to the NodeB, thus allowing the MAC-hs scheduler to extract multi-user diversity within the cell. Current schedulers typically are based on round robin or proportional-fair concepts [7]. Fig. 1 shows the main three elements of HSDPA networks, the UE, the NodeB—incorporating the MAC-hs scheduler—and the Radio Network Controller (RNC). Data flow control is handled by the MAC-d, ensuring continuous data transfer even in case of handover. These mechanisms together make the physical layer of HSDPA very flexible and promising for exploiting cross-layer information.

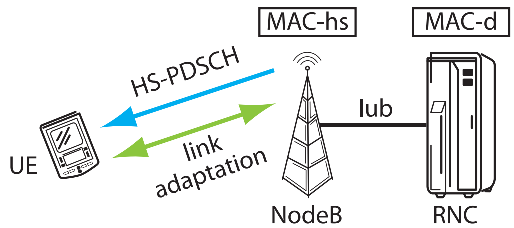

Fig. 1. Main physical layer associated network elements of HSDPA. The fast link adaptation (TTI based) is terminated at the NodeB.

Fig. 2 illustrates a more detailed look into the MAC-hs. MAC-d, MAC-hs and the physical layer set-up are controlled by the Radio Resource Control (RRC), which can adapt the network slowly to general performance targets. The available instantaneous Channel Quality Information (CQI) and the packet acknowledging (ACK/NACK) however are only available in the NodeB itself. The scheduler thus can act as an interface for the signaling information of upper layers and the CQI and ACK/NACK information of the wireless link. It has to be noted that, in principle, the scheduler may also influence the resources assignment of the individual users by affecting the modulation and coding settings as well as the retransmission handling.

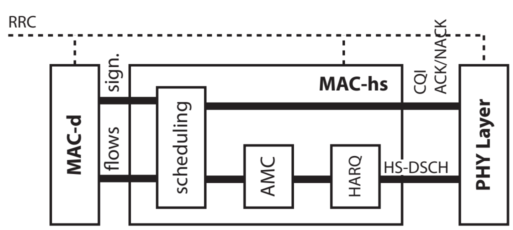

Fig. 2. Overview of the layer interaction responsible for the physical layer transmissions in HSDPA.

III. Video streaming over cellular Networks

Wireless video streaming is one of the most promising applications offered by the third generation mobile networks. Video streaming consists of the unicast transmission of an encoded video stream from a video server to an end user. In order to match the network capabilities and the terminals display sizes, the video is encoded with low resolution and low data rate. In this work we will consider the transmission of video encoded using the *state-of-art* video codec, the H.264/AVC [9] in its baseline profile, as recommended in [10].

Video encoding basically consists of reducing the correlation of the video sequence by exploiting its temporal and spatial redundancy. The baseline profile of the H.264/AVC allows two type of frames, the Intra (I) and the Inter (P) encoded ones. Each picture, or frame, is subdivided into blocks of 16×16 pixels, the so called MacroBlocks (MBs). In Intra encoded pictures, a spatial prediction of each macroblock is generated considering the already encoded neighboring MBs. Inter encoded macroblocks exploit the temporal correlation of consecutive frames: the best prediction of every macroblock is searched in the already encoded pictures. In both cases, the prediction is then refined by means of quantized frequency coefficients, called residuals. At the decoder, the inverse procedure is performed for reconstructing the picture. The intra encoded pictures are reconstructed using solely the code associated to the encoded frame, whereas the inter encoded pictures exploit also the already decoded pictures.

For transmission purposes, the Network Abstraction Layer (NAL) of the H.264/AVC segments the encoded stream into NAL Units (NALU). Each NALU is further encapsulated into Real Time Protocol (RTP), Universal Datagram Protocol (UDP) and Internet Protocol (IP). Since the size of an IP packet is limited by the network's Maximum Transfer Unit (MTU), each NALU contains a variable number of encoded macroblocks, referred as a picture *slice*.

Due to transmission errors as well as network congestion, the decoder has to implement strategies for handling missing NALUs. In case some packets are lost, the decoder reconstructs the picture slice using error concealment routines, i.e. by means of refined spatial and temporal interpolation. The period of intra frame encoding, defined as Group of Picture (GOP), has a direct impact on the quality of the video in an error prone transmission environment. Consider now the scenario depicted in Fig. 3, where one packet associated to $F_2^{(P)}$ is not correctly received. The missing information will affect the reconstruction of the frame $F_2^{(P)}$ as well as the frame $F_3^{(P)}$, that uses $F_2^{(P)}$ as source of prediction. The next I frame stops the temporal error propagation, since its reconstruction does not depend on the previously decoded frame

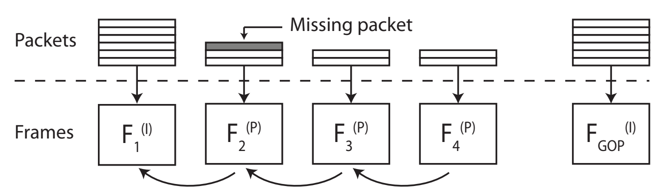

Fig. 3. Temporal dependency of encoded frames

Reducing the distance between consecutive I frames would, therefore, also reduce the extent of the temporal error propagation. However, since the spatial prediction is less effective than the temporal one, the size of the encoded I frames is larger than the one of the P frames. Thus, decreasing the GOP size causes the reduction of the coding efficiency.

IV. Quality of Service in UMTS Networks

The 3GPP already defined specifications for QoS in UMTS in [11], [12]. In UMTS, the QoS is obtained by setting up an end-to-end (human-human or server-human) bearer service with specified characteristics and functionalities. Beside the control signaling and the user plane transport, the service bearer attributes also comprise, among others, maximum and guaranteed bitrate, traffic handling priority and transfer delay.

Four classes of traffic have been defined: (i) Conversational: suitable for VoIP and videoconferencing tools, (ii) Streaming: covers the one-way application from a server (video or audio) to a human end user, (iii) Interactive: comprises human interaction with a remote server, such as web browsing, (iv) Background: collects the background applications where the end-user is not expecting data within a certain time.

At the same time, the IETF also defines its own architecture for providing QoS, the Differentiated Service (DiffServ) [13] architecture. The DiffServ, together with the Explicit Congestion Notification (ECN), inherited the byte of the IPv4 header before specified for defining the Type of Service (ToS) [14]. The five bits dedicated to the DiffServ are used to classify each data packet into a limited number of traffic classes. Beside the classification and DiffServ Code Point (DSCP) marking, DiffServ does not define how the traffic classes have to be handled by the network elements, this task is left to the network operator itself.

DiffServ does not establish a flow end-to-end quality of service, but rather specifies how a single packet with a given DSCP marking has to be handled by each network element in terms of queuing, scheduling and dropping mechanisms. Therefore, different Per Hop Behavior (PHB) classes has been defined: Default PHB, Assured Forwarding (AF) PHB, Expedited Forwarding (EF) PHB. To avoid long-term queuing while allowing short term congestion, DiffServ is usually implemented together with active queue management algorithms, such as Random Early Drop (RED) [15].

A set of communicating network elements agreeing on a common DiffServ specification is defined as DiffServ Domain (DD). An edge router is placed at the interface between the external network and the DiffServ domain. The edge router is responsible for cancelling any previous DCSP marking present in the packets accessing the network and for assigning them a new DSCP marking depending on the application and user class.

In the following, a brief overview of the literature discussing the possibility to implement QoS by means of DiffServ in UMTS networks will be offered. Several scientific publications [16], [17], [18], [19], [20] discuss the best strategy how to map UMTS QoS classes to DiffServ traffic classes. The promising results show how an appropriate mapping privileging applications, such as VoIP and video streaming, might be beneficial. However, these contributions just consider the definition of QoS on the core network and UTRAN network elements (in [17] the traffic bottleneck has been assumed to be between the GGSN and the SGSN). The characteristics of the wireless link between NodeB and user terminal has not been taken into consideration. In [21], the problem of differentiating packets within a single transmission flow is addressed. In a video telephony scenario, the authors discuss the possibility of handling differently the packets containing audio from the ones containing video, because of the respective traffic characterization and impact on the QoE. Still, the results are presented in terms of weighted delay and jitter.

V. Content-Aware Cross-Layer Optimization

In the following, we propose a practical implementation of a content-aware cross-layer optimization of H.264/AVC video streaming over UMTS networks.

Video streaming is an UMTS service usually belonging to the streaming QoS class. The video packets, flowing from a media server to the user terminal, are handled by the network elements in such a way that the time relation between them has to be possibly preserved. In Section III, it has been shown how I frames terminate temporal error propagation. The correctness of the packet containing I frames has a higher impact on the QoE. Assigning a single QoS class to the whole data flow, does not allow the network elements to better protect the most important payloads. In our contribution, we propose the implementation of a standard compliant signalling of a packet's priority, depending on its content, from the application layer to the MAC-hs layer.

At the application layer, H.264/AVC already allows the specification of the priority of a NALU in two reserved bits of the NAL header. This information can be easily conveyed through RTP and UDP to the IP header when generating the IP packets. In the following we describe a method to make use of this signaling information within the UMTS network. This is a non trivial task as the IP packets carrying the user data are encoded into other transport protocols between GGSN and mobile terminal, e.g., GTP (GPRS Tunnelling Protocol) and NBAP (Node B Application Part), see Fig. 4. The data itself is only available at the borders, i.e., GGSN and mobile terminal, of the UMTS network. However, there is a different and standard compliant way to use dynamic QoS settings for different packets.

In a UMTS network a logical packet switched connection is setup by a PDP-context. Within this method the user terminal at one side and the GGSN at the other side agree on several parameters, e.g., IP-address, PDP type and QoS profile for the following packets. In the UMTS network QoS only exists in this PDP-context entity, allowing for one setting at a time. To allow the system to serve independent QoS parameters for multiple applications running at the same mobile terminal a feature called multiple PDP-contexts was introduced.

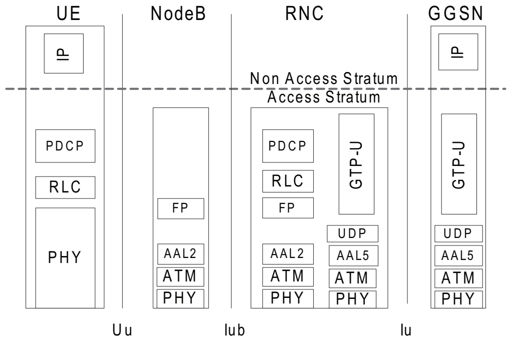

Fig. 4. UMTS Protocol Stack (User Plane).

There exist two different categories of PDP-contexts, namely primary and secondary. Every mobile has first to activate a primary context. Then the mobile can either attach a secondary to the primary or initiate further primary ones. Multiple primary PDP-contexts have different IP addresses and are typically used for different applications, e.g., email push services in parallel to standard Internet access. Multiple secondary PDP-contexts share the same IP address and must be attached to a previously initialized primary context. Figure 5 depicts the difference in the setup. As the IP address is the same for the primary and all the associated secondary PDP-contexts, a so called TFT (Traffic Flow Template) defines a split of user data into the GTP tunnels.

A TFT can be seen as a packet filter applied onto each IP packet entering the GGSN. There exists one TFT per secondary PDP-context. The filtering rules are based on one or more of the following attributes: source address, IP protocol number, destination port (range), source port (range), IPsec security parameter index and the type of service. The last attribute in this list makes the TFT the perfect match for the proposed method as it is the same header field as used for DiffServ.

We propose the use of multiple secondary PDP-contexts to transport the video data through the UMTS network with different settings of QoS according the information found within the packets. This solution is presented in Figure 5.

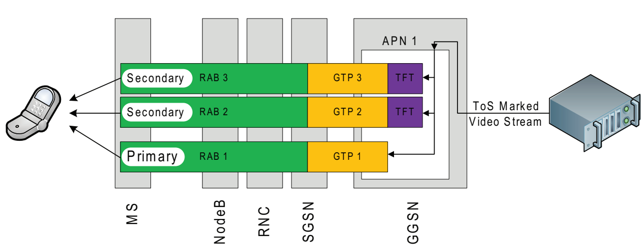

Fig. 5. Routing of QoS oriented data originating from one application server.

VI. Implementation and Results

In this section we discuss the improvements brought by a basic implementation of the proposed scheme. A video sequence in CIF resolution (352×288 pixels) has been encoded matching a target data rate of 250 kbit/s, the frame rate has been set to 15 frame/s. Different GOP sizes have been considered varying from 30 to 60 frames, respectively 2 to 4 seconds. Even doubling the GOP size, the resulting data rate decreases by only 3.69%.

The packets produced by the video encoder are subdivided into two flows: a primary PDP context contains the I frames, a secondary PDP context the P frames.

Our proposed method suggests protecting packets of higher priority better than those of lower priority. For a proof of concept of this idea, we have implemented a content-aware scheduler in the MAC-hs layer of HSDPA. The scheduler is built on a round robin implementation, but taking the priority information of the arriving IP-packets into account. Depending on the priority of the IP packets that fit into the transport block of the scheduled user, the MAC-hs scheduler sets the transmission settings accordingly. In particular this means that high priority transport blocks will be transmitted with a better protection against errors on the wireless link.

This procedure can be represented in terms of a so-called CQI-remapping, which applies the transmission settings associated to the CQI, c_{new}, that has been derived from the feedback CQI of the UE, c_{UE}, as given by

$$c_{\text{new}} = \Phi_p \left(c_{\text{UE}} \right), \tag{1}$$

where $\Phi_p(\cdot)$ denotes the remapping function as a function of the transport block priority p. By setting the priority of P frames to 1 and the priority of I frames to 2, for this work we simply set the remapping to be

$$\Phi_1(x) = x, \Phi_2(x) = x - 1, \tag{2}$$

thus being one-to-one for P frames and with a constant offset for I frames. Using a more conservative mapping for the I frames might cause the scheduler to assign to the considered user a smaller transport block size and, possibly, a less efficient modulation scheme. Since the standard transport block size and modulation scheme couples match a target BLock Error Rate (BLER), the conservative mapping is intended to reduce the BLER for the TBs containing Intra encoded frames. As a drawback, the throughput of the cell is slightly reduced, since the amount of data transmitted in a given TTI is smaller.

The implications of this scheduler implying this remapping have been tested by utilizing an HSDPA system level simulator [4], [5]. The simulation settings are given in Table I.

The results of the proposed method have been evaluated considering the overall quality of the video sequence using the standard round robin and the content-aware scheduling. The quality of the video has been measured in terms of Luminance Peak Signal to Noise Ratio (Y-PSNR). The Y-PSNR measures the error between the original frame (before encoding and transmission) and the frame reconstructed at the receiver.

As shown in Fig. 6, by means of content-aware scheduling the quality of the I frames has been increased, in average, of over 1 dB. As mentioned in Section III, the I frames stop the temporal error propagation. The correctness of the I frames, therefore, strongly influences the overall quality of

TABLE I
BASIC SIMULATION PARAMETERS FOR THE HSDPA SYSTEM-LEVEL EVALUATION OF THE CONTENT-AWARE SCHEDULER.

Parameter	Value
network deployment	19 cells, 3 sectors each
NodeB distance	1000 m
environment set-up	urban micro
channel type	ITU-T PedA
active users in target sector	5
UE capability class	10
UE receiver type	MMSE
UE speed	3 km/h

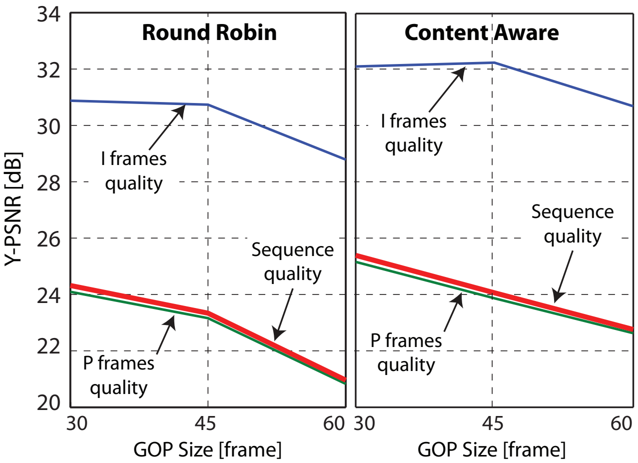

Fig. 6. Comparison of Round Robin and Content Aware Scheduling.

the whole sequence. It is worth noticing that, although the proposed mechanism is intended to improve the quality of the I frames, it is also beneficial for the P frames. If an I frame is incorrectly received, the quality of the following P frames, although correctly received, is affected.

VII. CONCLUSION

In this article we propose a scheme for improving quality of experience for video streaming over HSDPA networks. The logical connection between the video server and the user terminal has been subdivided into multiple PDP contexts, each with a specified QoS. Depending on the impact on the QoE, signalled by means of DiffServ marking, each video packet is associated to the appropriate PDP context. The payload of the most important flow is preserved by means of a more conservative CQI mapping. Simulations performed over an HSDPA system level simulator confirm the effectiveness of the proposed approach.

Further investigations are intended to analyze the behaviour of the method's components not observable by simulations. The establishment of the secondary PDP context might cost, in real systems, several seconds. This problem interests a small fraction of the average video session length and might be however overcome by transmitting the first frames using the primary connection and marking the packets addressed to the secondary context after an appropriate amount of time.

ACKNOWLEDGMENT

The authors thank mobilkom austria AG for technical and financial support of this work. The views expressed in this paper are those of the authors and do not necessarily reflect the views within mobilkom austria AG.

REFERENCES

[1] P. Svoboda, F. Ricciato, R. Pilz, E. Hasenleithner, "Composition of GPRS, UMTS traffic: snapshots from a live network", IPS-MOME 2006, Salzburg, Austria, Feb. 2006.
[2] H. Holma and A. Toskala, "WCDMA for UMTS – Radio Access For Third Generation Mobile Communications," 3rd ed. John Wiley & Sons, Ltd, 2005.
[3] H. Holma and A. Toskala, "HSDPA/HSUPA for UMTS – High Speed Radio Access for Mobile Communications", John Wiley & Sons, Ltd, 2006.
[4] M. Wrulich and S. Eder and I. Viering and M. Rupp, "Efficient Link-to-System Level Model for MIMO HSDPA," Proc. IEEE 4th Broadband Wireless Access Workshop, 2008.
[5] M. Wrulich and M. Rupp, "Computationally Efficient MIMO HSDPA System-Level Evaluation," EURASIP Journal on Wireless Communications and Networking, 2009, submitted.
[6] 3GPP, "Physical layer procedures (FDD)" (Tech. Spec. 25.214 V7.4.0), 2008. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/25214.htm
[7] T.E. Kolding, "Link and system performance aspects of proportional fair packet scheduling in WCDMA/HSDPA," IEEE Proc. Vehicular Technology Conference, September, pp. 1717–1722.
[8] 3GPP, "IP Multimedia Subsystem (IMS); Stage 2 (Tech. Spec. 23.228 V8.0.0), 2008. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/23228.htm
[9] ITU-T Rec. H.264 / ISO/IEC 11496-10, "Advanced Video Coding," Final Committee Draft, Document JVTE022, Sept. 2002.
[10] 3GPP, "Transparent end-to-end Packet-switched Streaming Service (PSS); Protocol and codecs," (Tech. Spec. 26.234 V7.3.0), 2007. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/26234.htm
[11] 3GPP, "Quality of Service (QoS) concept and architecture (Tech. Spec. 23.107 V8.0.0), 2008. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/23107.htm
[12] 3GPP, "Quality of Service (QoS) concept and architecture (Tech. Spec. 23.207 V8.0.0), 2008. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/23207.htm
[13] IETF, "An Architecture for Differentiated Service," (RFC 2475) 1998. [Online]. Available: www.ietf.org/rfc/rfc2475.txt
[14] IETF, "Type of Service in the Internet Protocol Suite," (RFC 1349) 1998. [Online]. Available: www.ietf.org/rfc/rfc1349.txt
[15] S. Floyd and V. Jacobson, "Random Early Detection gateways for Congestion Avoidance," IEEE/ACM Transactions on Networking, vol .1, no. 4, pp. 397-413, Aug. 1993
[16] F. Agharebparast, V.C.M. Leung, "QoS support in the UMTS/GPRS backbone network using DiffServ," IEEE GLOBECOM 2002, vol.2, pp. 1440–1444, Nov. 2002
[17] N.V. Lopes, M.J. Nicolau, A. Santos, "Efficiency of PRI and WRR DiffServ Scheduling Mechanisms for Real-Time Services on UMTS Environment," IEEE NTMS 2008. , pp.1–5, Nov. 2008
[18] E. Natalizio, S. Marano, A. Molinaro, "Packet scheduling algorithms for providing QoS on UMTS downlink shared channels," IEEE VTC 2005 (Fall), vol.4, pp. 2597–2601, Sept. 2005
[19] L.C. Saud, R.P. Lemos, "Third generation mobile wireless networks quality of service, with a 2.5G case study using Differentiated Services," IEEE Symposium on Advances in Wired and Wireless Communication, pp. 71–74, Apr. 2004
[20] H. Wang, D. Prasad, O. Teyeb, H.P. Schwefel, "Performance Enhancements of UMTS Networks Using End-to-End QoS Provisioning," IWS 2005, Sep. 2005
[21] R.B. Ali, S. Pierre, and Y. Lemieux, "UMTS-to-IP QoS mapping for voice and video telephony services," IEEE Network, vol. 19, nr. 2, pp. 26–32, March-April 2005.