Differentiated Services for Mobile Core Networks

Mirza Shahrukh Baig, Vishwanath Sinha*, Peter Stuckmann***

* Dept. of Electrical Engineering
Indian Institute of Technology Kanpur, India
vsinha@iitk.ernet.in, Mirza.Shahrukh@eed.ericsson.se

** Chair of Communication Networks
Technical University of Aachen, Germany
pst@comnets.rwth-aachen.de

Abstract

The mobile telecommunication community is migrating towards a new network architecture. While second-generation (2G) mobile core networks were originally circuit-switched, they will be based on the Internet Protocol (IP) with the introduction of the General Packet Radio Service (GPRS) into GSM networks. The Internet Protocol (IP) is not sufficient to serve traffic with specific latency, variance, packet loss, and throughput requirements. As a result, proposals are being made to improve the best-effort services of IP. Differentiated Services (DiffServ) is one such proposal, which is at present seen as the future technology for mobile core networks and other Internet networks when the focus is set on the scalability of network resources. This paper examines the feasibility of quality of service provisioning in mobile radio networks with the support of DiffServ in the core network compared to present IP technology realizing a pure best-effort service. As the radio interface GPRS is regarded as the first mobile packet technology based on IP infrastructure.

Keywords

DiffServ, Quality of Service, IP, General Packet Radio Service.

1. Introduction

In the framework of the evolution of the Global System for Mobile Communication (GSM) towards third-generation (3G) mobile communication systems known as the Universal Mobile Telecommunication System (UMTS) new radio interface standards are presently being integrated into existing mobile radio networks. After High Speed Circuit Switched Data (HSCSD) has been introduced in some countries, the General Packet Radio Service (GPRS) will be available in the year 2001 in Europe and many countries world wide. The next in line are EDGE enhancements of GPRS and the integration of new air interfaces like WCDMA realizing UMTS Terrestrial Radio Access Networks (UTRANs) [1, 2].

For the interconnection of these new radio access technologies with the information infrastructure, e.g., the public Internet, all these radio access technologies will be based on the same core network architecture. Core networks standardized by ETSI/3GPP are composed of IP routers, which realize the tunneling of user data to gateway IP routers and the interworking functions with subnetworks like external networks or other Public Land Mobile Networks (PLMNs).

While in GPRS phase one only best-effort data services without differentiating the applications will be supported, in the second phase quality of service (QoS) management functions will be integrated to be able to guarantee application-specific QoS requirements. These QoS functions will be based on the aggregation of flows belonging to the same service class and the prioritized admission control and scheduling of these aggregated flows in the radio network. No strict resource reservation is performed. This is the motivation to use DiffServ in GPRS/EDGE networks, since DiffServ is also a scalable approach without need for per-flow state and signalling at every hop, in contrast to the Integrated Services (IntServ) approach, which is based on the Resource Reservation Protocol (RSVP) [3].

In this paper we have compared the bahavior of a DiffServ-based and a best effort IP core network regarding Interactive (WWW) and Background applications (e-mail). To achieve this, an IP core network model is developed and integrated into a GPRS simulation environment. Then performance analysis is carried out considering both the radio network and the core network.

The paper is organized as follows: in Section 2 the DiffServ architecture and its relation to QoS functions in the radio network are presented. Section 3 describes the simulation model GPRSim to evaluate the performance of DiffServ compared to best-effort IP. In Section 4 the results are presented. Section 5 concludes this paper.

2. Differentiated Services

2.1. Network Architecture

Differentiated Services (DiffServ) is an approach to provide quality of service (QoS) within the Internet in a scalable manner. It is a relatively simple and coarse method of providing differentiated classes of service for various application and subscriber requirements.

The DiffServ architecture comprises a small, well-defined set of building blocks from which a variety of aggregate behaviors may be built. The QoS support may be end-to-end or intra-domain. Both quantitative performance requirements (e.g., peak throughput) and requirements based on relative performance (e.g. class differentiation) are supported.

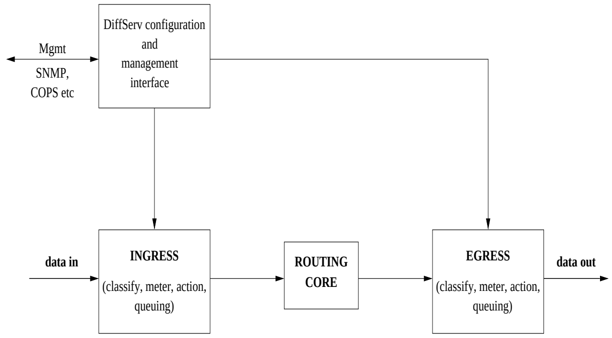

Figure 1: Differentiated Services architecture [4]

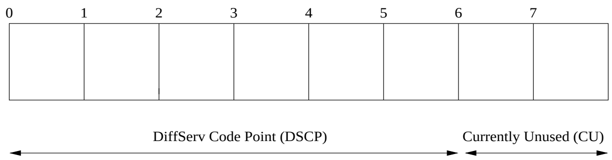

Figure 2: IPv4 Type of Service(TOS) octet or IPv6 Traffic Class octet

2.2. DiffServ Field

A small bit-pattern in each packet, in the IPv4 Type of Service (TOS) octet or the IPv6 Traffic Class octet, is used to mark a packet to receive a particular forwarding treatment, or per-hop behavior, at each network node. A common understanding of the use and interpretation of this bit-pattern is required for inter-domain use, multi-vendor interoperability, and consistent reasoning about expected aggregate behaviors in a network. Thus, the DiffServ Working Group (IETF) has standardized a common layout for a six-bit field of both octets, called the DiffServ field (DS field) [5].

In DiffServ, the network allows applications to negotiate one of the several different services through connection admission control (CAC) and the packets generated by applications, are treated differently by the network. DiffServ is expected to be used predominantly in IP backbone environments. With proper engineering, including boundary policing, DiffServ can provide expedited handling appropriate for a wide class of applications, including lower delay for mission-critical applications and packet voice applications. DiffServ-capable routers need to track a small number of per-hop behaviors, and they serve packets based on a single byte. Typically, DiffServ is associated with a course level of packet classification based on the DS field.

The architecture consists of a set of necessary functional elements:

Per-Hop Behaviors at Interior Routers: An interior router is any router not at the boundary of a DiffServ network domain. Since interior routers make up the vast majority of routers through which most IP packets pass, the complexity of the functions performed by interior routers must remain low. The DiffServ architecture recognizes this fact and mandates that only simple per-hop behaviors (PHBs) are implemented at interior routers. A per-hop behavior is any forwarding behavior performed by the router and usually consists of packet queuing and scheduling.

0,1,2 bits of DSCP	Precedence Level	
111	precedence 7	used for link layer and routing protocols
110	precedence 6	used for IP routing protocols
101	precedence 5	Express forwarding (expedited forwarding)
100	precedence 4	Assured forwarding class 4
011	precedence 3	Assured forwarding class 3
010	precedence 2	Assured forwarding class 2
001	precedence 1	Assured forwarding class 1
000	precedence 0	Best effort class

Figure 3: Precedence Levels of DiffServ based on bits 0, 1, 2 of DSCP

Traffic Classification and Conditioning at Boundary Routers: The boundary router is located at the edge of a DiffServ network domain. This boundary router must perform sophisticated packet classification, metering, marking, policing, and shaping operations.

2.3. Bandwidth Broker

To make appropriate internal and external admission control decisions and to configure boundary devices correctly, each DiffServ domain is outfitted with a bandwidth broker. The bandwidth broker performs admission control depending on networks resources and configures boundary routers to take or drop a connection request, e.g., it might request information from the queues of all routers on one or several paths before admitting a request.

As mentioned above the DiffServ architecture is based on the use of the DS Field which is placed in the IPv4 Type of Service octet or IPv6 Traffic Class octet (see Figure 2) [5]. The first six bits of the DS field are used as a codepoint (DSCP). These determine the per-hop behavior (PHB) a packet experiences at each node. PHBs define how traffic belonging to a particular behavior aggregate is treated at an individual network node. A two-bit currently unused (CU) field is reserved for future use. Depending upon the first three bits of DSCP we have 8 precedence levels (classes) in DiffServ as shown in Figure 3.

3. Interworking with the GPRS Radio Network

3.1. QoS Management in the GPRS Radio Network

3GPP has specified QoS requirements for different service classes namely Conversational, Streaming, Interactive and Background. Table 1 gives an overview of these requirements [6, 7].

To define a QoS contract between the mobile station (MS) and the network, Packet Data Protocol (PDP) contexts contain-

Table 1: QoS requirements for selected services belonging to different traffic classes

Traffic class	Medium	Application	Data rate	One-way delay
Conversational	Audio	Telephony	4–25 kbit/s	< 150 ms
	Data	Telnet	< 8 kbit/s	< 250 ms
Streaming	Audio	Streaming audio (HQ)	32–128 kbit/s	< 10 s
	Video	One-way	32–384 kbit/s	< 10 s
	Data	FTP	–	< 10 s
Interactive	Audio	Voice messaging	4–13 kbit/s	< 1 s
	Data	Web-browsing (HTML)	–	< 4 s/page

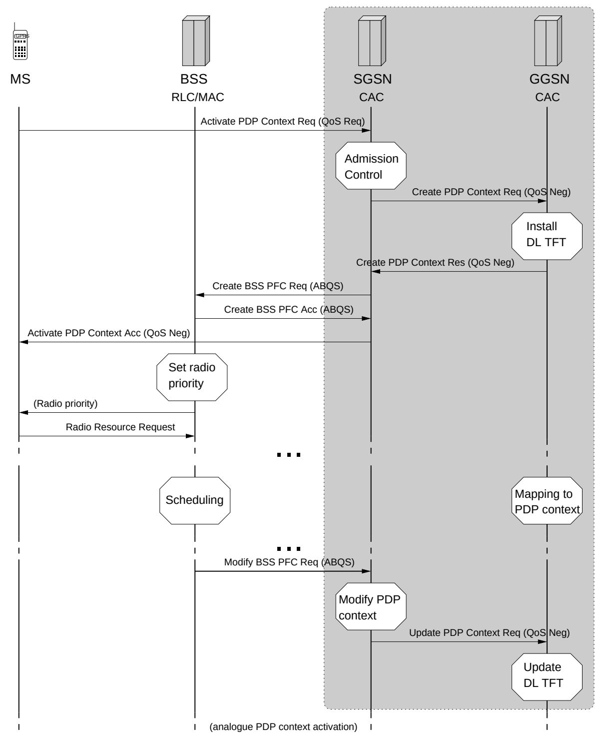

Figure 4: QoS negotiation and renegotiation procedures (example)

UMTS QoS class	DiffServ class	Reason
Conversational class	Expedited Forwarding class	As it requires low latency and jitter and EF class guarantees a minimun service level
Streaming class	Assured Forwarding class 4	As it requires low variation of delay i.e. stringent jitter requirements
Interactive class	Assured Forwarding class 3	As it requires low latency (but not as low as in conversational class)
Background class	Assured Forwarding class 2 or class1 or best effort (class 0)	As there is no specific requrement for this class except reliability, it can be given to AF class2 or AF class1 or even to best effort service

Figure 5: A proposal for mapping UMTS classes onto DiffServ classes

ing QoS profiles are negotiated between the MS and the Serving GPRS Support Node (SGSN) [8]. The Base Station Subsystem (BSS) is provided with a Packet Flow Context (PFC) containing the Aggregate BSS QoS Profile (ABQP) and is responsible for resource allocation on a Temporary Block Flow (TBF) base and scheduling of packet data traffic with respect to the according QoS profiles negotiated. Moreover, it regularly informs the SGSN about the current load conditions in the radio cell. The tasks of the Gateway GPRS Support Node (GGSN) comprise mapping of PDP addresses as well as classification of incoming traffic from external networks regarding the downlink Traffic Flow Template (TFT). The GPRS Register (GR) holds the QoS-related subscriber information and delivers it on demand to the GSN. In Figure 4 a GPRS session is schematically outlined depicting the instances involved, messages exchanged, and parameters used for PDP context (re)negotiation, PFC setup, and TFT installation [9].

3.2. A Poposal for Mapping GPRS/UMTS Classes onto DiffServ Classes

To integrate the QoS management functions in the radio network and the QoS functions in the IP core network, e.g. DiffServ, mapping rules have to be defined so that these functions can interwork efficiently. A proposal for mapping GPRS/UMTS classes onto DiffServ classes can be seen from Figure 5. In this proposal the Conversational class is mapped onto the expedited forwarding class as it is a real-time application and requires both low delay and low delay jitter. The Streaming class is mapped onto assured forwarding class 4 as it requires stringent delay jitter requirements. The Interactive class is mapped onto the assured forwarding class 3 as it requires low latency but not as low as in the Conversational class. The Background class can be mapped onto any lower DiffServ class.

4. Simulation Model

The GPRSim (see Figure 6) comprises the modules MS, BS, SGSN, the transmission links, the load generators, session control modules, a graphical user interface, and a module for statistical evaluation [10].

The MS, BS, and SGSN modules contain the implementations of the respective protocol stacks. Transmission links are represented by simple error models. While the G_b interface is

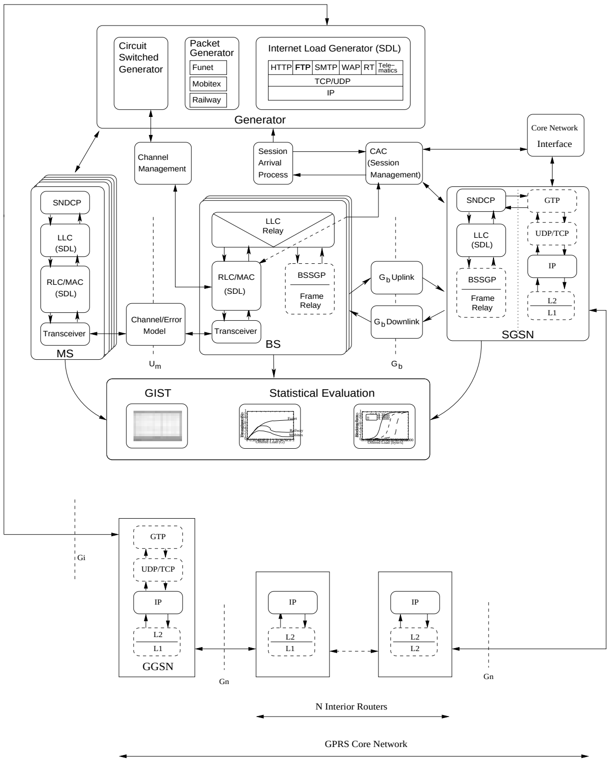

Figure 6: The GPRS Simulator GPRSim with the core network model

regarded as error-free, block errors on the radio interface, U_m are modeled based on look-up tables, which map a C/I value onto a block error probability (BLEP).

QoS functions in the radio network, namely Connection Admission Control (CAC) at the SGSN and scheduling in the Base Station Subsystem (BSS) are considered as described in [9].

An IP core network model consisting of boundary routers and a cascade of interior routers is developed and integrated into the GPRS simulator. All complex classification and conditioning is performed by these boundary routers with simple scheduling and queuing given to interior routers. The cascade model is suitable only as long as we are concerned with traffic on the same path between source and destination. Several different paths are not regarded. This model is sufficient to get statements about the capability of DiffServ to serve traffic even if routers in the core network are congested because of backbone traffic. Connection admission control is realized by a bandwidth broker sitting at the boundary routers, here the first SGSN connected to the BSS and the GGSN. It requests information of queues of all interior routers before admitting a request. In both boundary routers and interior routers the scheduling algorithm used is weighted round robin (WRR) scheduling. The reason for the choice is the DiffServ standard which says that the interior routers should be free from complex methods and work performed by interior routers should be a minimum. Any end-to-end connection will have a number of interior routers involved, hence its architecture should be simple. Another method of scheduling that can be used is weighted fair queuing (WFQ) but since it would require that each interior router should also know about the flow states, namely the number of flows passing through the router and their priority class, it is not regarded here.

Table 2: Model parameters of Internet applications (HTTP and SMTP)

HTTP Parameter	Distribution	Mean
Pages per session	geometric	5.0
Intervals between pages [s]	negative exponential	12.0
Objects per page	geometric	2.5
Object size [byte]	$\log_2$-Erlang-k	3700
Amount of SMTP data	**Distribution**	**Mean**
E-Mail Size [byte]	$\log_2$-normal	10000
Base quota [byte]	constant	300

5. Simulation Results

5.1. Simulation Scenarios

The GSM cell configuration is given by the number of transceiver units (TRX) in the radio cell. Here a typical 3-TRX scenario is regarded with 0 fixed and 8 on-demand Packet Data Channels (PDCH) that are shared with circuit switched GSM traffic, which is offered corresponding to an Erlang-blocking probability of 1 %. This means that on average around 7 PDCHs are available for GPRS [11].

The air interface error model is realized by a constant RLC/MAC block error probability of 13.5 % corresponding to a C/I of 12 dB. As the coding scheme CS-2 is used.

LLC and RLC/MAC are operating in acknowledged mode. The multislot capability is 1 uplink and 4 downlink slots. The MAC protocol instances in the simulation model are operating with three random access subchannels per 52-frame. LLC has a window size of 16 frames. TCP/IP header compression in SNDCP is performed. TCP is operating with a maximum congestion window size of 8 Kbyte and a TCP Maximum Segment Size (MSS) of 536 byte. The session interarrival time is set to 12 seconds. The Internet traffic [12] is composed of 70 % E-Mail sessions and 30 % WWW sessions (see Table 2) not depending on the subscription profile of the regarded MS.

The core network model is configured with an output rate for each router of 2 Mbit/s, and scheduling periods of 40 ms. Additional backbone traffic of 1 Mbit/s and 2 Mbit/s is generated to model a low load scenario and a congestion scenario in the core network.

For each background load scenario, one simulation series is performed with QoS management in the radio and core network (interactive and background) and one with a pure best-effort service without service differentiation in the radio and core network.

5.2. Performance and System Measures

As performance measures the average downlink IP throughput per user during a data transmission period and the average downlink IP packet delay are regarded. These are the QoS measures that are noticed by the user and that can be compared to the ETSI/3GPP QoS classes [6, 8].

For WWW and e-mail applications the throughput per user is the important measure since it mirrors the response time of a requested file.

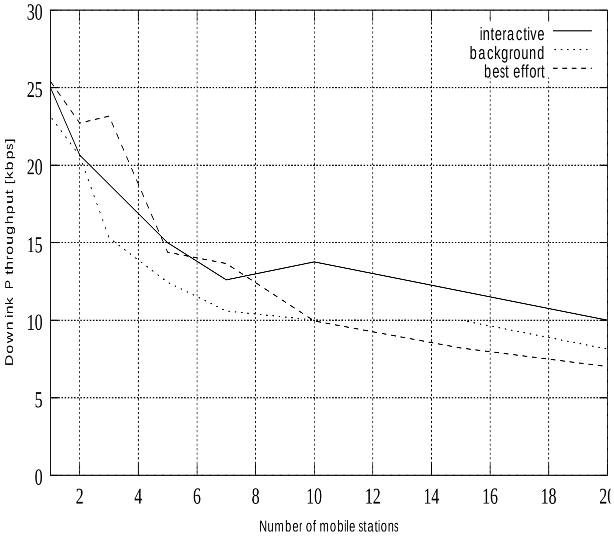

Figure 7: Downlink IP throughput per user with low backbone traffic load

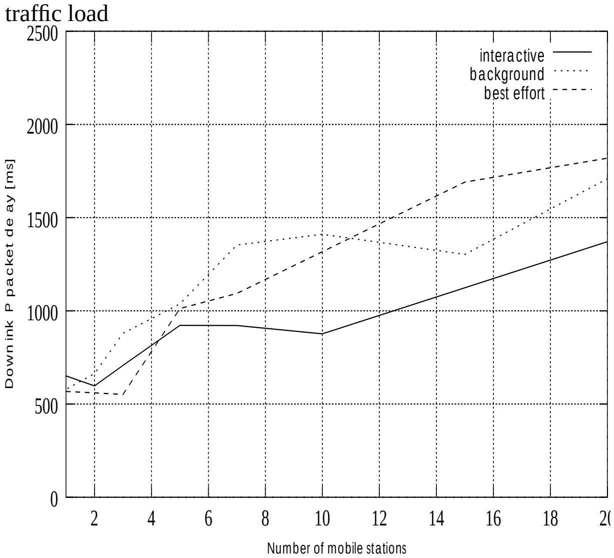

Figure 8: Downlink IP packet delay with low backbone traffic load

5.3. Simulation Results for Low Backbone Traffic Load

The Figures 7 and 8 show the performance measures throughput and delay for the Interactive (WWW) and Background (e-mail) classes, when service differentiation is done in the radio and core network. They are compared with the performance of the same traffic mix of WWW and e-mail, if no service differentiation in the radio and core network is done. Since the resources in the core network are not highly utilized by backbone traffic the core network can be seen as nearly transparent even for the best-effort GPRS traffic.

The difference in the performance of the different service classes can be explained by the QoS functions in the radio network in high radio network load situations [9]. A performance gain in throughput for Interactive applications compared to the best-effort service can be seen. For the Background class it is not enough to regard only the average throughput and delay measures, since more than 5 % of the Background sessions are

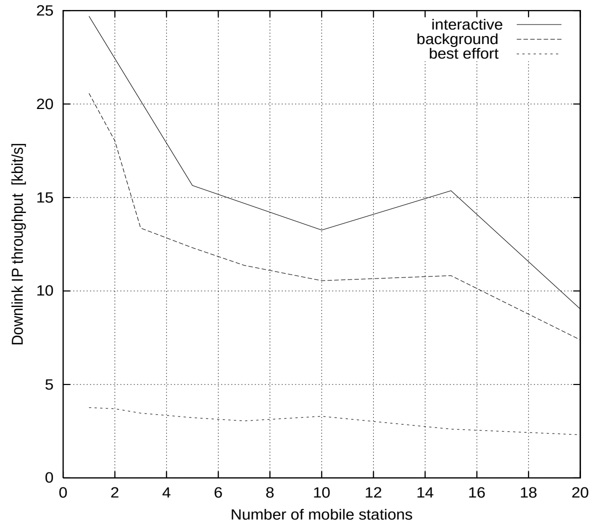

Figure 9: Downlink IP throughput per user with high backbone traffic load

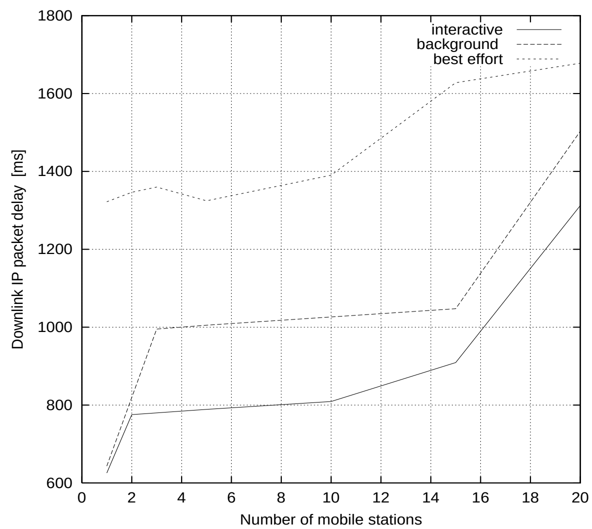

Figure 10: Downlink IP packet delay with high backbone traffic load

aborted in high load situations.

In this scenario with 2 Mbit/s router output rate and only 1 Mbit/s backbone traffic DiffServ does not have a significant influence on the performance, since there is enough capacity left for the regarded GPRS traffic. However, it can be seen that the prioritization of the Interactive class is supported by DiffServ and the results are nearly equal to the results presented in [9], where the influence of the core network is neglected.

5.4. Simulation Results for a Congested Core Network

More significant effects of DiffServ on the performance of different service classes become visible in a scenario with high backbone traffic load. In the regarded scenario the backbone traffic equals the output rate of the routers in the core network. Since the best-effort GPRS traffic will be served with the same priority as the backbone traffic, the congestion in the core network will effect the GPRS traffic. This results in poor performance for best-effort GPRS traffic. In Figure 9 and 10 this is shown with average IP throughput values below 5 kbit/s and average IP packet delay values above 1300 ms even in situations with low GPRS traffic load. If WWW and e-mail traffic is not served as best-effort traffic, but service differentiation both in the radio and the core network is performed, the same performance as in a low utilized core network can be reached through DiffServ QoS functions. The Figures 9 and 10 show that average downlink IP throughput values above 20 kbit/s and average downlink IP packet delay values below 800 ms are achieved in situations with low GPRS traffic load. With increasing GPRS traffic the performance decreases similarly as in the scenario without core network congestion (see Figures 7 and 8). This is only caused by the limited resources of 8 on-demand PDCHs in the GPRS radio network.

6. Conclusion

In this paper the capability of DiffServ to interwork with GPRS QoS functions is examined. To achieve this an IP core network model based on the DiffServ architecture, which is composed of two boundary routers and a cascade of interior routers, was developed and integrated into the GPRS simlation tool GPRSim.

With this model it has been shown that DiffServ is able to support service differentiation, which is done in the radio network based on GPRS/UMTS QoS classes, also in the core network. There is no difference to the performance of a scenario, where the core network influence is fully neglected.

If the core network is congested, a significant advantage compared to a best-effort service in the core network is achieved. While a congested core network without IP QoS functions would lead to poor performance for GPRS traffic, DiffServ is able to serve prioritized GPRS traffic similar to the performance without any influence of the core network.

As a result DiffServ is capable to interwork with GPRS QoS functions, so that the core network can be seen as transparent and nearly without influence on the GPRS traffic performance even in the congestion case.

7. References

[1] B. Walke, *Mobile Radio Networks - Networking, Protocols, and Traffic Performance*. Chichester: John Wiley & Sons, 2nd ed., to be published in 2001.

[2] G. Brasche and B. Walke, "Concepts, Services and Protocols of the New GSM Phase 2+ General Packet Radio Service," *IEEE Communications Magazine*, vol. 35, 1997.

[3] I. Costa and L. Dell'Uomo, "Quality of Service in Future 4G Mobile Wireless Systems," in *Proc. of the IEEE International Conference on Third Generation Wireless and Beyond (3Gwireless '01)*, June 2001.

[4] S. Blake, D. Black, M. Carlson, E. Davis, Z. Wang, and W. Weiss, "An Architecture for Differentiated Services, Request for Comments 2475," Technical Report, Internet Engineering Task Force (IETF), December 1998.

[5] K. Nichols, S. Blake, F. Baker, and D. Black, "Definition of the Differentiated Services Field in the IPv4 and IPv6 Headers, Request for Comments 2474," Technical Report, Internet Engineering Task Force (IETF), December 1998.

[6] ETSI 3GPP, "Universal Mobile Telecommunications System (UMTS); Service aspects; Services and Service Capabilities (3G TS 22.105 version 3.8.0 Release 1999)," Technical Specification ETSI TS 122 105, European Telecommunications Standards Institute, Sophia Antipolis, France, Mar. 2000.

[7] ETSI 3GPP, "Universal Mobile Telecommunications System (UMTS); QoS Concept and Architecture (3G TS 23.107 version 3.1.0 Release 1999)," Technical Specification ETSI TS 123 107, European Telecommunications Standards Institute, Sophia Antipolis, France, Jan. 2000.

[8] ETSI TC-SMG, "Digital cellular telecommunications system (phase 2+); general packet radio service (gprs); service description; stage 2 (gsm 03.60 version 7.4.0 release 1998)," Draft European Standard ETSI EN 301 344, European Telecommunications Standards Institute, Sophia Antipolis, France, April 2000.

[9] P. Stuckmann and F. Müller, "Quality of Service Management in GPRS Networks," in *Proc. of the IEEE International Conference on Networking (ICN '01)*, July 2001.

[10] P. Stuckmann, "Simulation Environment GPRSim: Tool for Performance Analysis, Capacity Planning and QoS Enhancement in GPRS/EDGE Networks," Technical Report, `http://www.comnets.rwth-aachen.de/~pst`.

[11] P. Stuckmann and F. Müller, "GPRS Radio Network Capacity Considering Coexisting Circuit-Switched Traffic Sources," in *Proc. of European Conference on Wireless Technology (ECWT 2000)*, 2000.

[12] P. Stuckmann and P. Seidenberg, "Quality of Service of Internet Applications over GPRS," in *Proc. of European Wireless 99*, vol. 1, 1999.