Comparative Analysis of Scheduling Algorithms for UMTS Traffic in Case of DiffServ Network Congestion

Sladjana Zoric

Faculty of Electrical Engineering and Computing, Zagreb, Croatia e-mail: sladana.zoric@gmail.com Mladen Kos Faculty of Electrical Engineering and Computing, Zagreb, Croatia e-mail: mladen.kos@fer.hr Himzo Bajric BH Telecom, Joint Stock Company Sarajevo, Bosnia and Herzegovina e-mail: hbajric@bhtelecom.ba

Abstract — Having in mind picture of future all-IP network, as well as fixed-mobile convergence, interworking between Universal Mobile Telecommunication System (UMTS) and external networks is crucial for providing end-to-end Quality of Service (QoS). For the purpose of achieving QoS in all-IP mobile networks, particularly for broadband multimedia services, Differentiated Services (DiffServ) mechanism should be applied to UMTS technology. This paper proposes using of Low Latency Queuing (LLQ) scheduler, with the main idea of mapping voice and video telephony in two different QoS classes and virtual queues, but at the same time using Priority Queuing (PQ) within LLQ for both voice and video telephony over all other QoS classes. To proof the concept, a simulation study was performed using Network Simulator version 2 (ns-2). Evaluation results of simulation study are presented for UMTS traffic that passes through UMTS core network and overloaded external IP backbone network. Performances of LLQ scheduler are compared with other most widely used scheduling algorithms such as Weighted Fair Queuing (WFQ), Weighted Round Robin (WRR) and Priority Queuing (PQ). The main objective of simulation was to provide QoS parameters such as IP Transfer Delay (IPTD), IP Delay Variation (IPDV) and IP Loss Rate (IPLR) for conversational and streaming traffic classes below standard defined values but not to completely exhaust bandwidth for interactive and background traffic classes. The obtained results were statistically processed using Statistical Package for the Social Sciences (SPSS) version 17.0.

Keywords - QoS; DiffServ; Scheduling; LLQ.

I. INTRODUCTION

Within network of Universal Mobile core Telecommunication System (UMTS), IP Multimedia Subsystem (IMS) [1] presents foundation of network which is completely based on Internet Protocol (IP) and provides support for multimedia services like Voice over IP (VoIP) and Video Streaming. As new multimedia services require higher restrictions in network parameters and have different requirements, the support for Quality of Service (QoS) is necessary. Prominent advantage of UMTS is its ability to provide diverse services with QoS guarantees. This paper will focus on the analysis of ensuring QoS for UMTS realtime traffic (Conversational and Streaming traffic class) in a mixed network environment, composed of the UMTS core network and IP external domain. For providing QoS in IP networks, IETF has developed different QoS mechanisms,

like Differentiated Services (DiffServ) [2] and Integrated Services (IntServ) [3]. IntServ have scalability and complexity problems, while DiffServ can be implemented in UMTS network with little or no management complexity. DiffServ architecture is based on a simple model where traffic entering a network is classified and conditioned at the boundaries of the network according to the Differentatied Services Code Point (DSCP) field in IP header, and assigned to different behavior aggregates. Within the core of the network, packets are forwarded according to the Per-Hop Behaviours (PHB) associated with the DSCP. PHB definitions do not specify any particular implementation mechanism and therefore the problem of PHB implementation has recently gained significant attention. According to 3GPP specifications, mapping of UMTS traffic classes into PHB can be done in gateway GPRS Support Node (GGSN), in order to get efficient PHB configurations [4]. Standard QoS mapping authorizes both voice and video telephony to be mapped to the same QoS class. Video traffic has larger packet sizes than voice traffic and can cause significant delay of voice packets when aggregating both together to the same QoS class. Authors in paper [5] have analyzed refined mapping between voice and video telephony but do not take other UMTS traffic classes into account. On the other side, in paper [6] are discussed QoS aspects both for real-time and non real-time traffic in UMTS simulation environment, but only in case of Priority Queuing (PQ) and Weighted Round Robin (WRR) schedulers. Dekeris, et al., [7] combine WFQ and LLQ, but the main drawback of this idea is the property that delay of high priority class (Video conferencing) could be reduced, but at the same time Voice traffic got the highest delay time. In our paper is presented concept of mapping voice and video telephony to different QoS classes and idea of implementing Low Latency Queuing (LLQ) traffic scheduler on network elements. Within LLQ, Priority Queuing (PQ) is used for scheduling of both voice and video telephony with respect to other traffic classes. Performances of LLQ scheduler are compared with other most available traffic scheduling algorithms such as Weighted Fair Queuing (WFQ), Weighted Round Robin (WRR) and Priority Queuing (PQ). In this paper, network congestion effect on QoS parameters for real-time traffic is investigated, that is conversational and streaming traffic classes. The aim of our proposed model was to provide IP Transfer Delay (IPTD), IP Delay Variation (IPDV) and IP Loss Rate (IPLR) to be below standard defined values, even in case of high network overload.

This paper is organized as follows. Section II makes a brief presentation of UMTS QoS model, the most commonly used PHB and our suggestion of mapping between UMTS and DiffServ domain. Section III provides overview of scheduling algorithms which will be used in simulation study. Section IV presents ns2 simulation model and simulation results, together with their discussion and analysis to show the conclusions that are warranted. Section V concludes this paper and describes direction for the future work.

II. UMTS TO DIFFSERV QOS MAPPING

3GPP standard proposes a layered architecture for the support of end-to-end QoS. To realize a certain network QoS, a Bearer Service (BS) with clearly defined functionalities has to be set up from the source to the destination of a service and includes all aspects to enable the provision of a contracted QoS.

UMTS BS attributes form a QoS profile and define the grade of service provided by the UMTS network to the user of the UMTS bearer service. UMTS specification [8] defines four traffic classes and they are: conversational, streaming, interactive and background. The main difference between these classes is how delay sensitive the traffic is. Applications of conversational and streaming classes are the most delay sensitive and intended for real-time traffic, while applications of interactive and background classes require higher reliability. Examples of applications are voice and video telephony for Conversational class and Video Streaming for the Streaming class. Interactive class is used by interactive applications like interactive web browsing, while Background class can be used for background download of e-mails.

Since the UMTS packet switched core network is based on an IP, DiffServ can be used for QoS provisioning. Figure 1 shows the example of how end-to-end QoS may be accomplished for a significant number of scenarios. In this paper, first scenario from 3GPP specification has been chosen, where the GGSN supports DiffServ Edge function and the IP network is DiffServ enabled. The application layer identifies QoS requirements, which are mapped into Packet Data Protocol (PDP) context parameters in UE. Local mechanism in the UE uses the PDP context for QoS over the UMTS access network, and the IP backbone network uses DiffServ to provide QoS guarantees. According to [5] IP BS manager is located in GGSN and uses standard IP mechanisms to manage IP bearer service. Provision of IP BS manager is optional in User Equipment (UE) and mandatory in the GGSN. Translation/Mapping function in GGSN provides interworking between the mechanisms and parameters used within the UMTS bearer service and those used within the IP bearer service. It is operator's choice to define the mechanisms for the provisioning of resources among the different DiffServ PHB classes, as well as the mapping from the UMTS QoS classes, to the DSCP. The DiffServ working group of IETF has defined different PHB groups for different applications.



Figure 1. Network architecture for QoS conceptual model [5]

The EF-PHB [9] is intended to support low-loss, lowdelay and low-jitter services. The EF guarantees that traffic is serviced at a rate that is at least equal to a configurable minimum service rate (regardless of the offered load and non-EF traffic) at both long and short intervals. IETF defines AF-PHB group in [10]. AF allows the operator to provide assurance of delivery as long as the traffic does not exceed some subscribed rate. Traffic that exceeds the subscription rate faces a higher probability of being dropped if congestion occurs. AF PHB defines four independent PHB classes, each with three dropping precedence level. Each corresponding PHB is known as AF_{ii} , where *i* represents AF class, while *j* is the drop precedence. Within an AF class, packets of drop precedence p experience a level of loss lower (or equal to) than the level of loss experienced by packets of drop precedence q if p < q. Each AF class is configured with separate buffer and bandwidth. Default PHB has best effort forwarding characteristics.

Service provider should consider all the UMTS QoS classes that are defined in network, aggregate these classes into a manageable set of new groups, based on their QoS requirements. Correspondingly, a set of available PHB that have similar characteristics should be chosen, and a one-toone mapping assigned. Traditionally, all traffic in Conversational class, i.e. voice and video telephony, should be mapped to the same EF class which is intended for critical voice traffic. In fact, voice packets have short and constant packet size while video packets have large and variable packet size. When injecting voice telephony traffic together with bursty video telephony traffic, video traffic can cause degradation as well as delay of voice service. Therefore we suggest the mapping of voice and video telephony to two different DiffServ virtual queues and DiffServ classes. Voice telephony is mapped to EF class and video telephony to AF11 class. Aggregate of Streaming class is mapped into AF21, as it requires low variation of delay and has higher delay constraint than the Interactive class but less constraint than the Conversational class Aggregate of Interactive class is mapped to AF31, and do not have special requirements, except reliability. Aggregate of Background class is mapped to default PHB.

TRAFFIC CLASS	PHB	DSCP VALUE
Conversational voice	EF	46
Conversational video	AF11	10
Streaming video	AF21	18
Interactive	AF31	26
Background	BE	0

TABLE I. QOS MAPPING TABLE

III. TRAFFIC SCHEDULING ALGORITHMS

PHB simply characterizes the externally observable forwarding behavior of a DiffServ router to the corresponding traffic stream. PHB definitions do not specify any particular implementation mechanism. To instantiate a particular PHB, network administrator activates and tunes an appropriate combination of specific packet-scheduling algorithms and Active Queue Management (AQM) mechanisms supported by the DiffServ router. The choice of a traffic scheduling algorithm is important for the implementation of behavior aggregates in a DiffServ network. When multiple queues are sharing common transmission media, there must be a scheduler to decide how to pick up packets from each queue to send out and is responsible for enforcing resource allocation to individual flows. If there is no congestion on the interface, packets are transmitted as they arrive. If the interface is experiencing congestion, scheduling algorithms are engaged. Scheduler performances have the highest impact on the level of service a packet receives [11]. The most popular and available scheduling algorithms in IP routers, and used in our simulation are: WFQ, WFF, PQ and LLQ.

A. Priority queuing (PQ)

In classic PQ, packets are first classified by the system and then placed into different priority queues. Packets are scheduled from the head of the given queue only if all queues of higher priority are empty. Within each of the priority queues, packets are scheduled in FIFO order. Benefit of PQ is relatively low computational load on the system. The biggest problem of using PQ is if the volume of higherpriority traffic becomes excessive, lower priority traffic can be dropped as the buffer space allocated to low-priority queues starts to overflow.

B. Weighted Round Robin (WRR)

In WRR, packets are first classified into various service classes and then assigned to a queue that is specifically dedicated to that service class. Each of the queues is then serviced in a round robin (RR) order. The weight indicates how many packets have to be sent in each cycle from each queue. The WRR scheduler doesn't take the size of the transmitted packets into account. As a result, it is difficult to predict the actual bandwidth that each queue obtains, but it ensures that all service classes have access to at least some configured amount of network bandwidth.

C. Weighted Fair Queuing (WFQ)

WFQ supports flows with different bandwidth requirements by giving each queue a weight that assigns it a different percentage of output port bandwidth. WFQ supports the fair distribution of bandwidth for variablelength packets by approximating a generalized processor sharing (GPS) system. GPS [12] assumes that the input traffic is infinitely divisible and that all sessions can be served at the same time. GPS is a theoretical model and in reality it cannot be implemented. WFQ schedules packets according to their arrival time, size, and the associated weight. Upon the arrival of a new packet, a "virtual finish time" is calculated which represents time at which the same packet would finish to be served in the GPS system. WFQ outputs packets in the ascending order of the virtual finish time.

D. Low Latency Queuing (LLQ)

LLQ is a combination of PQ and Class-Based Weighted-Fair Queuing (CBWFQ). CBWFQ extends the standard WFQ functionality to provide support for user-defined traffic classes. The LLQ, like PQ checks the low-latency queue first and takes a packet from that queue. If there are no packets in the low-latency queue, the normal scheduler logic applies to the other non-low-latency queues, giving them their guaranteed bandwidth. LLQ allows delay-sensitive applications such as voice to be given preferential treatment over all other traffic classes [13].

IV. SIMULATION MODEL

Simulation is performed using network simulator ns2, which is an event-driven simulator targeted at networking research [14] and independent developed module for scheduling algorithms used in this paper [15]. Default implementation of LLQ within ns2 simulator, which supports scheduling of only one queue with PQ is changed in order to support scheduling of two queues with PQ (for voice and video telephony). The aim of simulation is to evaluate performances of our proposed idea in terms of QoS parameters and perform comparison with other schedulers such as WFQ, WRR and PQ where voice and video telephony are mapped to the same virtual queue and traffic class. According to ITU-T Recommendation Y.1541 [16], QoS parameters for conversational traffic class should be:

- IP Transfer Delay, IPTD $\leq 100 \text{ ms}$,
- IP Delay Variation, IPDV \leq 50 ms,
- IP Loss Rate, IPLR $\leq 10^{-3}$,

while for streaming traffic class these parameters should be:

- IP Transfer Delay, IPTD \leq 400 ms,
- IP Delay Variation, IPDV \leq 50 ms,
- IP Loss Rate, IPLR $\leq 10^{-3}$.

Simulation model is presented in Figure 2. UMTS infrastructure is not fully simulated (radio interface), only the core network between SGSN and GGSN which is not congested. This is not inconsistent with the concept of UMTS architecture which specifies that access and core networks are independent [17].



Figure 2. Simulation model

The capacities of links in external IP backbone are dimensioned in a way to implement network configuration whose load equals 20%. Starting from this configuration, we decrease the capacity of appropriate links gradually to 200% according to the amount of traffic passing through external IP backbone, which is 6Mb/s and is constant during all simulation. Voice telephony traffic sources are represented with Exponential (EXP) traffic generator, which generates 4 flows with packet size of 80 bytes and 100 kb/s rate, while video telephony sources generate 5 flows with 1000 bytes and 100kb/s rate. Constant Bit Rate (CBR) generator is used in order to generate 16 video traffic flows with packet size of 1000 bytes and 300 kb/s rate. Interactive traffic sources are simulated with Telnet application, and generate 12 flows with 500 bytes packets size with 100 kb/s rate. Background data sources are configured with File Transfer Protocol (FTP) traffic generator with 1000 bytes packet length and 10 kb/s rate from 18 flows. CBR and EXP traffic generators are attached to UDP agents, while FTP and Telnet traffic generators are attached to TCP agents. Conversational traffic (voice and video telephony together) produces 15%, streaming traffic produces 80%, while interactive and background traffic produce 2% and 3% of overall generated traffic. Time Sliding Window 2 Color Marker (TSW2CM) is used as a policer to determine how to mark and prioritize the packet according to user requirements. Weighted Random Early Detection (WRED) is used as Active Queue Management mechanism for Streaming and Interactive traffic, while for Conversational and Background traffic is used Drop Tail. In case that PQ scheduler is used, Conversational traffic has the highest priority, while Background traffic has the lowest priority. Weights of other schedulers are configured in such a way, that weight represents percentage of output port bandwidth: 15 for Conversational, 80 for Streaming, 2 for Interactive and 3 for Background traffic class. The queue lengths are constant and are defined with 30 packets for Conversational class and 50 packets for all other traffic classes. Simulation results in this paper are depicted only for second flow generated from all traffic sources, which is chosen randomly, but could be for any of the generated flows.

As we can see from Figures 3a and 3b, average end-toend delay for Conversational traffic class (for both voice and video telephony) stays within 100 ms only when PQ and LLQ are implemented on network nodes. Considering the effect of different schedulers on average end-to-end delay for Streaming traffic, which is depicted in Figure 3c, we can notice that all schedulers have almost the same performances and provide satisfactory level of QoS according to reference [16]. Results from Figures 3d and 3e show that jitter stays below 30 ms for voice telephony and below 35 ms for video telephony in all experiments. The same behavior is also observed in Figure 3f, for Streaming traffic, where jitter is lower than 12ms for all schedulers. Results for jitter show non-monotonic behavior: increasing with the network load, reaching some maximum and the decreasing. More network latency is necessary in order to deliver a stream due to network congestion. In Figures 3g and 3h packet loss rate for voice and video telephony is presented. Network congestion has the greatest influence on WRR scheduler, which does not perform satisfactory when network is overloaded more than 100%. On the other side, there is no packet loss for PQ and LLQ schedulers. Similarly, in Figure 3i, packet loss rate for Streaming traffic is presented; we see that it remains below 0.2% for all schedulers except WRR, even under heavy network congestion of 200%.

Table II depicts the average values of the link throughput between nodes CORE2 and EDGE for Interactive and Background traffic classes. As expected, when PQ scheduling algorithm is used, lower priority classes are starving and throughput is equal zero for both Interactive and Background traffic classes when network congestion in higher than 120%. LLQ on the other side provides fair level of bandwidth for lower priority classes, and at the same time fulfills QoS requirements for Conversational and Streaming traffic classes. From all these experiments, it can be concluded that our approach of using LLQ scheduler and mapping of voice and video into two different classes



Figure 3. Simulation results: (a) End-to-end delay of voice telephony; (b) End-to-end delay of video telephony; (c) End-to-end delay of video streaming; (d) Jitter of voice telephony; (e) Jitter of video telephony; (f) Jitter of video streaming; (g) Packet loss rate of voice telephony; (h) Packet loss rate of video telephony; (i) Packet loss rate of video streaming

provides critical performance parameters for voice and video telephony below standard defined values. These results are consequence of handling Conversational traffic with strict priority over all other traffic classes.

The obtained results are statistically processed using Statistical Package for the Social Sciences (SPSS) version 17.0. The null hypothesis states that QoS parameters can be guaranteed when using LLQ. Against the null hypothesis is setup the alternative hypothesis. The 95% confidence interval is chosen, which relates to level of statistical significance of p<0.05. The regression analysis is conducted to find the relationship that explains how the variation in IPTD/IPDV/IPLR values for Conversational and Streaming traffic classes, depends on the variation in network overload. Coefficient of correlation (r) is measured to give the true direction the correlation, while the coefficient of

determination (r^2) is measured to give the strength of correlation. Using comparative analysis of different regression models in SPSS, we have decided to use cubic polynomial regression model since it has the highest value of determination coefficient r^2 . The results of regression analysis have been summarized in Table III only for our approach of using LLQ scheduler as it provides the best performances among all other simulated schedulers.

 TABLE II.
 Average Throughput For Interactive and Backgound Traffic Classes

TRAFFIC CLASS	WFQ	WRR	PQ	LLQ
Interactive	8.93	7.79	4.62	8.71
Background	37.20	33.07	27.40	35.41

	Regression model	b ₀	b 1	b ₂	b ₃	\mathbf{r}^2	r	р
Voice telephony IPTD	$y=b_0+b_1t+b_2t^2+b_3t^3$	-325.36	0	-0.120	0.002	0.997	0.997	>0.05
Video telephony IPTD	$y=b_0+b_1t+b_2t^2+b_3t^3$	-934.68	0	0.138	0	0.999	0.999	>0.05
Video streaming IPTD	$y=b_0+b_1t+b_2t^2+b_3t^3$	-22.713	0.867	0	7.36*10 ⁻⁷	0.891	0.944	>0.05
Voice telephony IPDV	$y=b_0+b_1t+b_2t^2+b_3t^3$	367.128	-977.614	945.438	-248.369	0.781	0.609	>0.05
Video telephony IPDV	$y=b_0+b_1t+b_2t^2+b_3t^3$	-9.78	229.924	148.92	-200.7	0.941	0.885	>0.05
Video streaming IPDV	$y=b_0+b_1t+b_2t^2+b_3t^3$	26.883	141.885	-69.228	12.068	0.987	0.974	>0.05
Voice telephony IPLR	-	-	-	-	-	-	-	-
Video telephony IPLR	-	-	-	-	-	-	-	-
Video streaming IPLR	$y=b_0+b_1t+b_2t^2+b_3t^3$	59.96	484.788	2610.032	0	0.902	0.814	>0.05

TABLE III. STATISTICAL ANALYSIS OF QOS PARAMETERS FOR CONVERSATIONAL AND STREAMING TRAFFIC CLASSES - LLQ SCHEDULER

From Table III, we can see that regression dependency between packet loss and link load percentage for conversational traffic is not possible to determine. Value of packet loss rate is constant and is always zero for conversational traffic class no matter how high is traffic overload in external IP backbone network. For all other results depicted in Table III, p value (significance) is greater than 0.05, which means that there is almost no statistical relationship between QoS parameters and link load percentage when LLQ scheduler is used.

V. CONCLUSION AND FUTURE WORK

QoS as an end-to-end concept has to be satisfied through the interworking of all the entities the UMTS traffic is passing through. In order to achieve desired end-to-end performances, it is crucial to define efficient QoS mapping scheme between UMTS services and IP QoS classes in case of DiffServ based network. This paper presented one example of mapping which was implemented on GGSN and EDGE nodes, as they perform DiffServ edge function in our simulation model. That approach suggested mapping of voice and video telephony into two different QoS classes and virtual queues. Other important problem which was pointed out in our work concerns the implementation of PHB and the choice of traffic scheduling algorithm. We proposed the idea of using LLQ scheduler, with PQ scheduling for both voice and video telephony over all other traffic classes. Default implementation of LLQ scheduler in ns2 has been changed in order to support scheduling of two virtual queues with Priority Queuing. The results from our simulation study indicate that using LLQ provides better performances than using WFQ, PQ and WRR schedulers in terms of QoS parameters such as IPTD, IPDV and IPLR for real-time UMTS traffic. Results obtained from statistical analysis indicate that there is almost no statistical relationship between the performance metrics of Real-time services and the network load when novel approach is used. Future work will focus on hierarchical traffic scheduling in order to perform refined scheduling between voice and video telephony.

REFERENCES

- S. Zorić, J. Baraković, and H. Hodžić, "QoS Architecture in IP Multimedia Subsystem of UMTS", 50th International Symposium ELMAR-2008, Zadar, Croatia, September 2008.
- [2] S. Blake, et al., "An architecture for Differentiated Services", Technical Report RFC 2475, IETF, December 1998.
- [3] R. Braden, D.Clark and S.Shenker, "Integrated Services in the Internet Architecture: an Overview", Technical Report RFC 1633, IETF, June 1994.
- [4] TS 23.207 Release 7, "End-to-end Quality of Service, concept and architecture", 3GPP, June 2007.
- [5] R.B. Ali, S. Pierre, and Y. Lemieux, "UMTS-to-IP QoS mapping for voice and video telepfony services", IEEE Network, vol. 19, no. 2, pp. 26-32, 2005.
- [6] N.V.Lopes, M.J. Nicolau, and A. Santos, "Efficiency of PRI and WRR DiffServ Scheduling Mechanisms for Real-Time Services on UMTS Environment", NTMS, November 2008.
- [7] B. Dekeris, T. Adomkus, and A. Budnikas, "Analysis of QoS assurance using weighted fair queuing (WFQ) scheduling discipline with low latency queue (LLQ)", 28th Int. Cof. Information Technology Interfaces ITI, Cavtat, Croatia, June, 2006.
- [8] TS 23.107 Release 7, "Quality of Service, concept and architecture", 3GPP, September 2007.
- [9] V. Jacobson, et al., "An Expedited Forwarding PHB", Technical Report RFC 2598, IETF, June 1999.
- [10] J. Heinanen, et al., "Assured Forwarding PHB Group", Technical Report RFC 2597, IETF, June 1999.
- [11] S. Zoric and M. Bolic, "Fairness of Scheduling Algorithms for Real-Time Traffic in DiffServ Based Networks", 15th IEEE Mediterranean Electrotechnical Conference MELECON, Valletta, Malta, April 2010.
- [12] A. Parekh and R. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single Node Case", ACM/IEEE Transactions on Networking, pp. 344-357, June 1993.
- [13] S. Zoric, M. Kos, and H. Bajric, "Fairness of scheduling algorithms for Real-time UMTS traffic in case of IP Link Congestion", ICUMT 2011, Budapest, Hungary, October 2011.
- [14] Network Simulator, http://www.isi.edu/nsnam/ns, April 2012
- [15] http://sergioandreozzi.com/research/network/diffserv4ns/, April 2012
- [16] ITU-T Recommendation Y.1541, "Network performance objectives for IP-based services", February 2006.
- [17] S.I. Maniatis, E.G. Nikolouzou, and I.S. Venieris, "QoS issues in the converged 3G wireless and wired networks", IEEE Communications Magazine, Volume 40, Issue 8, 2002