 The definitive version of this paper has been published in Infocom, 2003, 10.1109\/infcom.2003.1208948.

MEDF - A Simple Scheduling Algorithm for Two Real-Time Transport Service Classes with Application in the UTRAN

Michael Menth
Institute of Computer Science
University of Würzburg, Germany
E-Mail: menth@informatik.uni-wuerzburg.de

Matthias Schmid
Infosim AG, Germany
E-Mail: schmid@infosim.net

Herbert Heiß and Thomas Reim
Siemens AG, Germany
E-Mail: Herbert.Heiss@siemens.com
E-Mail: Thomas.Reim@siemens.com

***Abstract*— In this paper, we consider real-time speech traffic, real-time circuit-switched data (CSD) and non-real-time packet-switched data (PSD) in the UMTS Terrestrial Radio Access Network (UTRAN). The focus is on the single low-bandwidth link that interconnects the Radio Network Controller (RNC) and the base station (Node B). We show that all traffic on this link has real-time requirements. But we take advantage of the Radio Link Control (RLC) layer protocol and formulate suitable quality of service (QoS) criteria that lead to two different transport service classes (TSC): A stringent TSC for speech traffic and CSD, and a tolerant TSC for PSD. The RNC transmits packets from both TSCs via a single low-bandwidth link to the Node B. Since transmission capacity on this interface is a serious cost factor in the UTRAN, the link utilization should be optimized while respecting the QoS requirements of both TSCs. We propose a modified version (MEDF) of the Earliest Deadline First (EDF) algorithm for that task. In contrast to EDF, the MEDF is easy to implement in hardware and in contrast to algorithms like Weighted Fair Queuing (WFQ), the knowledge of the traffic mix is not needed for a suitable parameter setting in the MEDF scheduler. The simulation results show its superiority over First-In-First-Out (FIFO), Static Priority (SP), and Weighted Round Robin (WRR) scheduling. The analysis of the waiting time distribution explains why MEDF performs better than the other scheduling strategies.**

I. Introduction

Third generation mobile systems like the Universal Mobile Telecommunication System (UMTS) are designed to provide a wide range of services and applications to the mobile user. The support of higher user bitrates is most likely the best known feature of UMTS. Furthermore, provisioning of appropriate quality of service (QoS) will be one of the key success factors for UMTS. A mobile user gets access to UMTS through the Wideband Code Division Multiple Access (WCDMA) based UMTS Terrestrial Radio Access Network (UTRAN). A base station (Node B) terminates the level 1 air interface and forwards the (uplink) traffic to the Radio Network Controller (RNC). The RNC is responsible for the radio resource management (RRM) and controls all the radio resources within its part of the UTRAN (cf. Figure 1). The RNC is the key interface partner for the mobile user equipment (UE) and it is the interfacing entity towards the core network (via a UMTS Mobile Switching Center (UMSC) or a Serving GPRS Support Node (SGSN)). Within the UTRAN, the Asynchronous Transfer Mode (ATM) is used as the main transport technology for terrestrial interconnection of the UTRAN nodes (RNC, Node B) [1]. The transport over IP technology is also foreseen for future releases but it is not within the scope of this document.

This work was partially funded by the Bundesministerium für Bildung und Forschung of the Federal Republic of Germany (Förderkennzeichen 01AK045). The authors alone are responsible for the content of the paper.

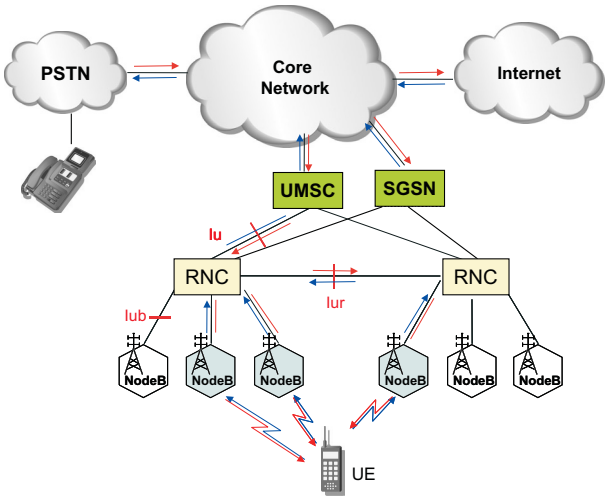

Figure 1. The architecture of the UMTS.

This paper focuses on the Iub interface, i.e. the interface between RNC and Node B. From a terrestrial transport network point of view this interface is the most critical one. On the one side this interface is overprovisioned to support QoS. On the other side, the bandwidth between RNC and a Node B becomes limited and expensive with the extension of the UTRAN to a more and more rural coverage. Therefore, Node B and RNC must be enabled to use the limited transmission resources to a maximum possible amount while meeting the QoS requirements.

This work is structured as follows. In Section 2, we describe the architecture in the UTRAN in more detail and point out

the requirements for the transport of user data. We define two different real-time transport service classes (TSCs) to differentiate the user traffic on the transmission link between RNC and Node B. In Section 3, we discuss various scheduling alternatives for the traffic handling on the Iub interface and propose the "Modified Earliest Deadline First" (MEDF) algorithm. Section 4 illustrates the performance of the considered scheduling variants and shows the superiority of the MEDF scheduler. Section 5 concludes this work with a short summary and an outlook on further research.

II. Transmission Requirements in the UTRAN

The 3^{rd} Generation Partnership Project (3GPP) has defined four different user traffic classes for UMTS [2]: the conversational, the streaming, the interactive and the background class. Conversational and streaming are mainly intended for real-time applications (voice, video, circuit-switched data) whereas the interactive and the background class are foreseen for Internet applications (e.g. e-mail, web browsing). In this paper we concentrate on three typical UMTS applications: Adaptive Multi-Rate (AMR) encoded voice (about 12.2 kbit/s), circuit switched data (CSD, 64 kbit/s), and packet-switched data (PSD, 64 kbit/s). CSD represents UMTS conversational and streaming services except for voice which is handled separately. PSD is a representative for interactive and background services. For the transport between the RNC and the UE, each user traffic flow is protected by the Radio Link Control (RLC) layer. It provides retransmission capabilities at least for the background and interactive services (PSD). This allows for the differentiation of traffic flows on Iub regarding their QoS requirements.

As outlined above, the RNC controls and assigns the radio resources. The Medium Access Control (MAC) in the RNC schedules the traffic to the UEs for the air interface depending on the user traffic class, radio frequency (RF) interference, power control efficiency, soft handover synchronization, and others. As a consequence of this sophisticated radio resource control, air interface resources are reserved for the traffic before its transmission at the Iub interface and cannot be used by other traffic. Data loss and excessive delay in the wireline transport network result in a waste of radio resources or capacity. Therefore, all traffic scheduled at the RNC for the air interface or received from the air interface has real-time QoS requirements.

For the purpose of air interface traffic scheduling, the MAC controls the characteristics of the traffic, i.e. transmission time interval (TTI) and the amount of user data to be sent. The TTI determines the time between two consecutive user data blocks that may be sent on the air interface. Each user traffic stream (voice, video, data) is carried by a so-called Dedicated Transport Channel (DCH) between RNC and Node B and the Frame Protocol (FP) provides for adaptation, error detection (optional), and exchange of control information for each DCH. The DCH traffic stream is handed over from the FP layer to the Transport Network Layer (TNL) in the form of FP Protocol Data Units (PDUs). Due to the MAC scheduling, the FP PDUs arrive periodically and a FP PDU carries the amount of user traffic to be sent on the air interface for one TTI (cf. Figure 2).

As mentioned before, ATM constitutes the current TNL technology in the UTRAN. The ATM Adaptation Layer Type 2 (AAL2) [3], [4] is applied to make the transmission of low-bitrate real-time traffic more efficient over ATM virtual channel connections (VCC) [5]. The AAL2 segments the FP PDUs into AAL2 packets with a CPS SDU (Common Part Sublayer Service Data Unit) of up to 45 octets and a header of 3 octets. Each DCH traffic stream is mapped to a periodic AAL2 packet stream that is transported on an AAL2 connection through the ATM network. CPS SDUs are not placed one-to-one into ATM cells but back-to-back as a continuous stream. This avoids unused payload in ATM cells if AAL2 packets are too small to complete the payload of an ATM cell (48 bytes). If AAL2 is applied, the first bit is used for signaling purposes, therefore, the effective payload for CPS SDUs of an ATM cell is 47 bytes. Up to 248 AAL2 connections can be multiplexed into an AAL2 path which corresponds to an ATM VCC.

The transmission links on the Iub are costly and limited in bandwidth. Low bandwidth induces considerable delay for the transport of the FP PDUs from RNC to Node B and vice versa. The delay may be further increased by the bursty nature of the AAL2 packet stream, especially if the TTI of a higher bitrate DCH is long [6]. Excessive delay of FP PDUs violates their real-time requirements and results in poor usage of the air interface resources. Therefore, the delay, which refers to the whole FP PDU transmission time and not only to a single AAL2 packet, must be bounded.

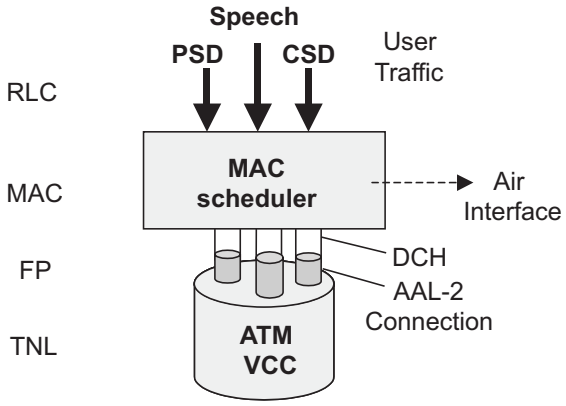

Figure 2. Iub traffic handling in the RNC.

On the one hand, FP packets demand for low latency on the Iub link. On the other hand, a high utilization of the costly transmission capacity is also an important objective in the UTRAN. We resolve this conflict by relaxing the delay requirements for the real-time transport on Iub. We do not guarantee a maximum delay for FP packets but expect that most of the traffic arrives at the Node B within a given delay budget (DB). We postulate this for 99.99% of the CSD and speech traffic. PSD can tolerate a higher packet loss probability because this traffic is protected by the RLC retransmission

capabilities, so the 99% quantile of their waiting time must not exceed the DB. Hence, we have two different real-time TCSs with stringent and tolerant QoS requirements. The objective for the multiplexing of AAL2 traffic on the Iub is to find a simple and intelligent scheduling algorithm that maximizes the utilization of the link while respecting the QoS requirements of the real-time traffic types [7], [8]. Currently, priority scheduling is implemented in the RNCs and only one priority is used, which is FIFO scheduling. Note that the actual problem is not ATM specific and occurs in IP networks, too.

III. Scheduling Alternatives on a Single Link

To accomplish the above stated task, we consider several well known scheduling algorithms and propose a modified version of EDF.

A. FIFO Scheduling

First-in-First-Out (FIFO) is probably the simplest queuing strategy: all packets are stored in a single queue in the order of their arrival and are served sequentially, regardless which QoS requirements they have. This is simple to implement but no service differentiation is possible and, therefore, no advantage can be taken from the lower QoS demand of tolerant traffic in order to increase the link utilization.

B. SP Scheduling

A simple queuing strategy for service differentiation is Static Priority (SP) scheduling. Packets are assigned different priorities and for every priority TSC a FIFO queue is set up. The queues with higher priorities are served exhaustively before queues with lower priority. This can lead to substantial queuing delay for the packets of the lower TSCs. In particular, the tail of the waiting time distribution can be non-exponential [9] because a majority of high priority traffic can delay low priority traffic extensively. Since this is not acceptable in the above sketched scenario, SP scheduling is not a good solution for the Iub interface.

C. WRR Scheduling

Generalized Processor Sharing (GPS) shares the capacity of a server according to given weights among different fluid queues. The traffic mix has to be known to set the weights of the scheduler in an appropriate way and a wrong parameter setting can have detrimental effects for the QoS of traffic. Weighted Fair Queuing (WFQ) [10] is the GPS version for packet oriented networks. In [11] a procedure is found to set the weights of WFQ to provide various queues with different bounded delay service. WFQ acts like a traffic spacer if low-bitrate streams are transported within a larger pipe, i.e. the packet delay is large if the data are bursty and if only the mean rate is allocated. Since WFQ is hard to implement in real-time [12], we chose Weighted Round Robin (WRR) as a simple approximation. Round Robin (RR) takes turns for servicing its queues and every queue receives the same share of bandwidth. With WRR, weights are assigned to the queues and some queues can be served more frequently by prescribing a serving cycle. The network capacity is shared among queues 0 and 1 with a ratio of 2:1 by using the serving cycle 0-0-1. If a queue has nothing to send, the next queue in the cycle is served.

D. EDF Scheduling

FP PDUs can afford a certain delay budget for the transmission over the Iub interface, i.e. they have a deadline (due date) of DB time after their arrival in the RNC. This calls for the Earliest Deadline First (EDF) algorithm [13] that schedules the packets according to ascending deadlines. It is an optimal strategy to maximize the throughput while meeting delay constraints [14]. However, hardware implementations of EDF are complex [15] because sorting is required. At the Iub interface, the delay budget is the same for all FPs and, therefore, the EDF scheduling order does not differentiate from FIFO. Hence, EDF does not provide any service differentiation among different TSCs.

E. MEDF - A Simple Algorithm for Service Differentiation

The "Modified Earliest Deadline First" (MEDF) algorithm is a modified version of the EDF. The packets are stored in n TSC specific queues in a FIFO manner. They are stamped with a modified deadline which is their arrival time plus an offset M_i, $0 \leq i < n$, that is characteristic for each TSC. The MEDF scheduler selects the packet for transmission that has the earliest due date among the packets in the front positions of all queues. For only two TSCs, this is the choice between two packets and sorting according to ascending deadlines is not required. Therefore, the MEDF scheduler is much simpler to implement in hardware than the EDF version. Furthermore, the modified deadline is not anymore an indicator for the latest arrival date of the packet. The difference $|M_i - M_j|$ between two TSCs i and j is a relative delay advantage that influences the behavior of the scheduler. This makes the MEDF suitable for the AAL2 mulitplexing on the Iub interface and distinguishes it from the conventional EDF. An analysis of a similar concept in the EDF scheduling context is given in [16].

F. Comparison of the Scheduling Alternatives

Essentially, the MEDF algorithm performs a local reordering within a FIFO queue. The maximum delay advantage between two queues determines the time interval in which reordering may occur, therefore, it must be smaller than the delay budget. We illustrate the behavior of the MEDF and for a better understanding we substitute the quantiles of the waiting time simply by the mean delay. In Figure 3 the packets carry their unique number in the upper part of their representation and their arrival time in the lower part. The light coloured packets have high priority (TSC 0) and the dark coloured packets have low priority (TSC 1). Both are scheduled onto a common link.

The scheduling discipline may be FIFO, SP, or MEDF scheduling and influences the produced packet schedule. Figure 4 shows the results for the proposed example. For FIFO

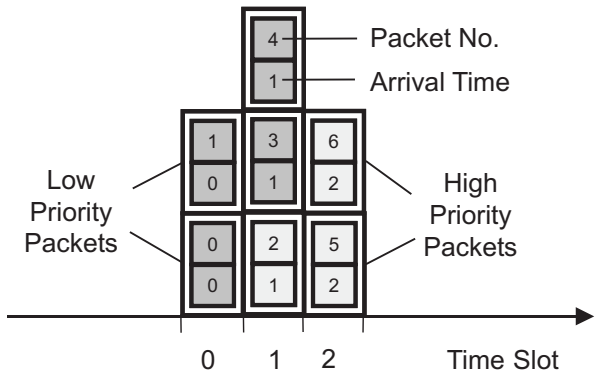

Figure 3. A packet arrival pattern in a time slotted system with two priority TSCs.

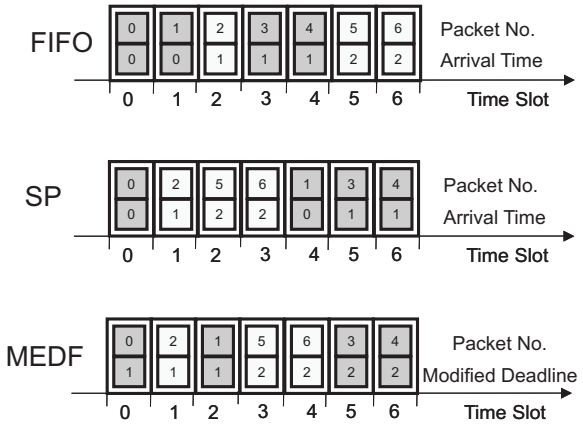

Figure 4. FIFO, SP, and MEDF scheduling produce different schedules.

and SP scheduling, the lower part in the packet representation tells again the arrival time of the packet. For MEDF scheduling we used $M_0 = 0$ and $M_1 = 1$ to set the modified deadline which is denoted in the lower part. From the time axis we can infer the transmission time of the packets.

TABLE I

MEAN DELAY IN TIME SLOTS.

	Low Priority Traffic	High Priority Traffic
FIFO Scheduling	1.50	2.67
SP Scheduling	3.25	0.33
MEDF Scheduling	2.75	1.00

This example leads to the mean delay statistics given in Table I. We assume a delay budget of 3 time slots. With FIFO scheduling the delay budget is met by the mean but in this example the mean delay for high priority traffic is substantially larger than for low priority traffic. SP scheduling expedites high priority packets at the expense of low priority packets whose mean delay exceeds the delay budget. With MEDF both low and high priority traffic fulfill the delay requirements and the delay for high priority traffic is smaller than for low priority traffic. This illustration is only a motivation to understand why MEDF scheduling is superior to FIFO and SP scheduling. In the next section we underline this statement by simulation results in the UTRAN.

IV. Performance Tradeoffs for Packet Scheduling in the UTRAN

In this section, we investigate the maximum link utilization between the RNC and the NodeB depending on the traffic mix and the scheduling algorithm in the RNC.

A. Simulation Setup

We simulated an AAL2 path (ATM VCC) from the RNC to the NodeB. The minimum transmission capacity on the Iub interface is typically $E1$ or 2 $E1$. The bandwidth in our simulations is 1.5 $E1$ (3.072 Mbit/s). We take three UMTS services into account: AMR speech (12.2 kbit/s), CSD and PSD (64 kbit/s). For all three data types we choose a TTI of 20 ms, i.e. FP packets are sent periodically every TTI. All the traffic types are periodic with the same TTI. Hence, the traffic mix has also the same period. As a consequence, the experienced waiting time distribution is almost the same in all TTIs as long as the traffic does not change. To obtain statistical data, the periodical arrival pattern of the packets within a TTI is carefully randomized.

FP packets from CSD and PSD are too large for AAL2 transport. They are segmented and sent in bursts of 4 CPS SDUs from the RNC to the Node B with altogether 176 and 184 bytes, respectively. For voice, we transformed representative speech samples [17] using an AMR vocoder [18] and captured the length distribution of the resulting CPS SDUs (cf. Table II). In addition, we also modeled the correlation in the size of successive speech frames of a single connection by a Memory Markov Chain [19]. Both are used for the simulation of an individual voice call. Signaling and other control and background traffic have only minor impact on the purpose of our paper and are not considered in our simulations.

TABLE II

FP PACKET LENGTH DISTRIBUTION FOR AMR VOICE ON THE IUB INTERFACE.

packet length [bytes]	probability
6	0.475
8	0.073
35	0.452

Like [6] we require the delay budget to be 5 ms. The critical load is the maximum link utilization for which the QoS criteria are still met. It is the performance objective in our investigation. A 3.072 Mbit/s link can carry at most 48 connections with 64 kbit/s which leads to a coarse granularity in the simulation results.

B. The Performance of FIFO Scheduling

The scheduling strategies have different service differentiation capabilities. Their impact is visible only for traffic mixes. For homogeneous traffic (only speech, CSD, or PSD traffic), the behavior of all mentioned scheduling alternatives becomes FIFO. The maximum link utilization is significantly higher for PSD traffic than for CSD traffic (cf. Figure 5)

because a smaller percentile of the traffic (99% for PSD instead of 99.99% for CSD) has to respect the delay budget. In contrast to CSD, voice packets do not arrive in bursts. The reduced burstiness induces shorter AAL2 multiplexer queues and decreases the delay for voice. Therefore, the critical load is higher for speech traffic than for PSD traffic (cf. Figure 6) in spite of the stricter delay criterion.

A FIFO scheduler handles all traffic in the same way and packets from both TSCs face the same delay statistics. Hence, the delay criterion for the stringent QoS traffic limits the maximum link utilization. As a consequence, the FIFO scheduler can not take advantage of large proportions of tolerant PSD that can afford an exceeded delay budget more frequently than stringent QoS traffic. The performance is only influenced by the queuing property of the traffic mix. CSD and PSD have almost the same burstiness, and the CSD:PSD traffic mix ratio has only little effect on the queuing properties of the traffic (cf. Figure 5). Speech traffic is less bursty than CSD, and PSD traffic and yields a larger link utilization (cf. traffic mix 5:0 in Figure 6) than pure CSD traffic. The performance of FIFO scheduling decreases with an increasing proportion of bursty PSD in the speech:PSD traffic mix. The queuing properties of the traffic mixes also influence the results for other scheduling strategies.

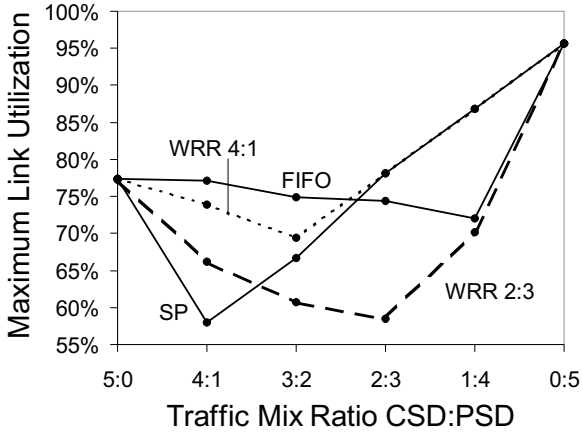

Figure 5. The performance of FIFO, SP and WRR scheduling for different weights and CSD:PSD traffic mixes.

C. The Performance of SP Scheduling

SP scheduling privileges high priority data (stringent QoS) at the expense of low priority data (tolerant QoS). For large proportions of PSD in the traffic mix, the critical load is higher for SP than for FIFO because the relaxed QoS criterion for tolerant traffic can be exploited. However, for large proportions of stringent QoS traffic in the traffic mix, the performance of SP is worse than for FIFO: The majority of stringent traffic delays the minority of tolerant traffic to such an extent that the relaxed QoS criterion for tolerant data can only be met for a low load.

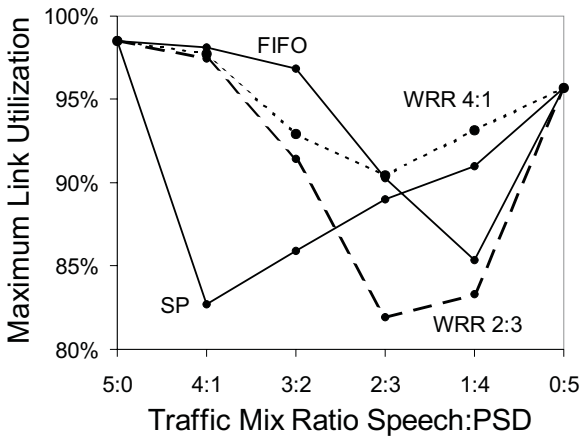

Figure 6. The performance of FIFO, SP and WRR scheduling for different weights and speech:PSD traffic mixes.

D. The Performance of WRR Scheduling

Suitable weights must be chosen for the WRR scheduler. A simple approach is to set them according to the traffic mix. This works fine for a traffic mix speech/CSD:PSD=4:1 but it yields even the worst performance for a traffic mix of speech/CSD:PSD=2:3 (cf. Figure 5 and Figure 6). Hence, the WRR scheduler is difficult to control. The weights 4:1 seems to be a good compromise for all traffic mixes considered but other weights (not given in the figures) yield a better performance for some specific traffic mixes. Hence, the information about the traffic mix must be propagated to the scheduler, otherwise it can not work in an optimal way. But if this information can be propagated, one can also apply a combination of FIFO and SP scheduling since there is no traffic mix for which WRR is significantly better than both of them.

E. The Performance of MEDF Scheduling

The setting of the offset parameters influence the MEDF algorithm. We choose $M_0 = 0$ for the stringent traffic and alter M_1 for tolerant traffic. It is obvious that MEDF yields the same results as FIFO scheduling for $M_1 = M_0$. For $M_1 \geq DB$, MEDF scheduling approximates SP scheduling: The modified deadline of a stringent packet is always smaller than the modified deadline of a concurrent tolerant packet provided that both arrive in time. In Figure 7 the MEDF seems to interpolate between FIFO and SP scheduling for different values of M_1 ($M_0 = 0$) for a traffic mix 4:1. For an offset parameter $M_1 = DB/4$, the MEDF scheduler is superior to the FIFO, SP, and WRR scheduling strategy (cf. Figures 7–8). MEDF scheduling respects the deadlines of the individual packets and the parameter $M_1 = DB/4$ effects that stringent packets are prioritized over tolerant packets but only within a short interval such that low priority traffic is not delayed for too long time. Note that the parameters M_i are chosen independently of the traffic mix which makes the MEDF easier to use than WFQ derivatives for which the optimal parameter settings depend on the traffic mix.

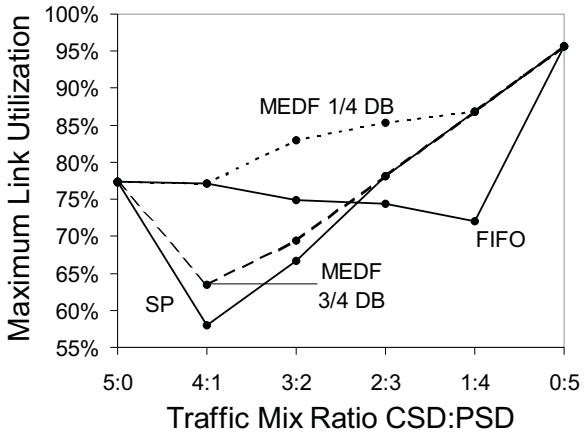

Figure 7. The performance of FIFO, SP and MEDF scheduling for different M_1 ($M_0 = 0$) and CSD:PSD traffic mixes.

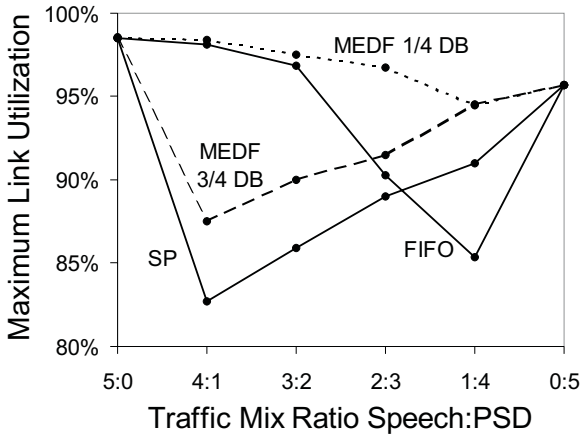

Figure 8. The performance of FIFO, SP and MEDF scheduling for different M_1 ($M_0 = 0$) and speech:PSD traffic mixes.

F. Scheduler Analysis Using the Waiting Time Distribution

We analyze the complementary distribution function (CDF) of the waiting time under critical load. In particular, we have a look at speech:PSD traffic mixes with a ratio of 1:4 and 4:1. The following observations validate and complete our previous explanations.

Figures 9–10 show the percentile of the traffic that faces a larger delay than a specified time. In case of MEDF scheduling, the delay budget of 5 ms is exceeded by the waiting time of PSD with a probability of less than 10^{-2} and for speech traffic with a probability of less than 10^{-4}. The MEDF algorithm distributes the overall queuing time to packets from both TSCs in a way that both meet their QoS criteria very closely under critical load. This is desirable but the other scheduling algorithms do not achieve that. In case of FIFO queuing, all packets are handled in the same way. For both TSCs, the 99.99% quantile of the waiting time distribution is smaller than the delay budget. For the sake of clarity, we omitted these curves in the figures. With SP scheduling, the waiting time of speech is significantly shorter than required because voice packets are prioritized over PSD. The QoS criterion for PSD limits the load on the link and the probabilistic QoS criterion of speech can not be entirely exploited. This finding also holds for WRR scheduling with weights 4:1 because this almost works like SP in the presence of a traffic mix speech:PSD=1:4 (cf. Figure 10). But WRR 4:1 does not strictly prioritize speech over PSD. Therefore, it achieves a larger critical load than SP (cf. Figure 6). In Figure 9 the CDF for speech lies above the CDF for PSD, although the weights correspond to the traffic mix. WRR is a packet oriented scheduling mechanism that is unaware of packet sizes (in contrast to the more complex WFQ). Connections with larger packets get a larger share of the bandwidth. Therefore, voice packets suffer a larger delay compared to PSD packets. If PSD packets suffer less delay than voice packets, vacant resources to increase the link utilization are wasted. This makes WRR even more difficult to use because besides the traffic mix the size of the packets must also be taken into account for appropriate parameter settings.

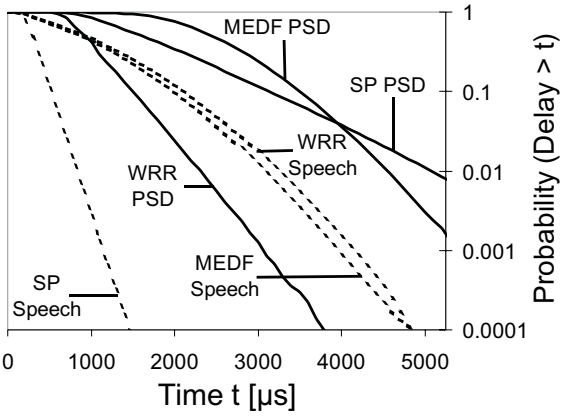

Figure 9. Optimal scheduling parameters and speech:PSD traffic mix 4:1.

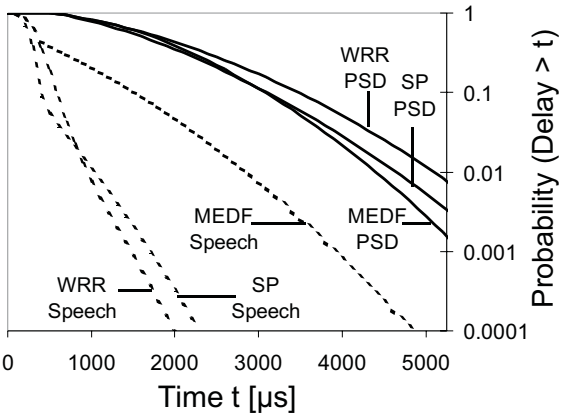

Figure 10. Optimal scheduling parameters and speech:PSD traffic mix 1:4.

In our study, the MEDF algorithm is the best scheduling strategy because it distributes the waiting time to packets from both TSCs such that the probabilistic QoS criteria are fully exploited under critical load. FIFO, SP, and WRR are not able to achieve that and, therefore, the utilization of the link is worse compared to MEDF.

V. Conclusion

We analyzed the transport requirements of speech, CSD and PSD services on the UMTS Iub interface. We formulated suitable QoS criteria and derived two different real-time transport service classes (TSCs): A stringent TSC for speech traffic and CSD and a tolerant TSC for PSD. Traffic of both TSCs is multiplexed onto an ATM VCC between RNC and Node B using AAL2. The goal was to maximize the link utilization on the Iub interface while meeting the QoS requirements of both TSCs by the use of an intelligent and simple scheduling strategy.

FIFO and EDF scheduling can not take advantage of the weaker QoS criterion for tolerant QoS traffic. SP scheduling suppresses low priority streams if the proportion of high priority traffic is large. Therefore, the load on the link must be kept small to respect the QoS criteria of both TSCs. WRR leads for some traffic mixes to a worse performance than SP and FIFO and can hardly increase the link utilization. In addition, the knowledge about traffic mix is required to set the optimal parameters for WRR scheduler. This information is usually not available.

MEDF, our modified EDF scheduler adaptation, stores packets in TSC specific queues in a FIFO manner. They are stamped with a modified deadline which is their arrival time plus a queue specific offset. The packet with the earliest due date among the packets in the front positions of all queues is eligible for transmission. This scheduler is easy to implement. We found a parameter setting such that the MEDF achieves an optimal link utilization independently of the traffic mix. Our simulations show that MEDF outperforms FIFO, SP, and WRR scheduling under all conditions. We further confirmed the superiority of MEDF scheduling by an analysis of the waiting time distribution.

The MEDF algorithm is not subject to ATM networks, only. We intend to further investigate its application in IP networks for the differentiation of various real-time and non-real-time TSCs in the Internet.

References

[1] G. Eneroth, G. Fodor, G. Leijonhufvud, A. Rácz, and I. Szabó, "Applying ATM/AAL2 as a Switching Technology in $3rd$ Generation Mobile Networks," *IEEE Communication Magazine*, vol. 37, June 1999.

[2] 3GPP, "3G TS 23.107 V3.6.0: QoS Concept and Architecture," Oct. 1999.

[3] ITU-T, "I.366.1 Segmentation and Reassembly: Service Specific Convergence Sublayer for the AAL Type 2," June 1998.

[4] ——, "I.366.2 Draft Recommendation: AAL Type 2 Service Specific Convergence Sublayer for Narrow-Band Services," June 2000.

[5] N. Gerlich and M. Menth, "The Performance of AAL-2 Carrying CDMA Voice Traffic," in *11th ITC Specialist Seminar*, Yokohama, Japan, Oct. 1998.

[6] S. Nananukul, Y. Guo, M. Holma, and S. Kekki, "Some Issues in Performance and Design of the ATM/AAL-2 Transport in the UTRAN," in *IEEE Wireless Communications and Networking Conference*, Sep. 2000.

[7] O. Isnard, J.-M. Calmel, A.-L. Beylot, and G. Pujolle, "Handling Traffic Classes at AAL2 / ATM Layer over the Logical Interfaces of the UMTS Terrestrial Access Network," in 11^{th} *IEEE International Symposium on Personal, Indoor and Mobile Radio Communication (PIMRC)*, London, England, Sept. 2000, pp. 1464–1468.

[8] J.-L. C. Wu, C.-H. Huang, and R.-T. Sheu, "Performance Study of AAL2 Protocol for Low-Bit-Rate Multimedia Services," in 15^{th} *International Conference on Information Networking (ICOIN)*, Beppu City, Japan, Jan. 2001, pp. 1464–1468.

[9] V. G. Subramanian and R. Srikant, "Tail Probabilities of Low-Priority Waiting Times and Queue Lengths in MAP/GI/1 Queues," *Queuing Systems*, vol. 34, no. 1-4, pp. 215–236, 2000.

[10] R. Guérin and V. Peris, "Quality-of-Service in Packet Networks: Basic Mechanisms and Directions," *Computer Networks*, vol. 31, no. 3, pp. 169–189, Feb. 1999.

[11] K. Kumaran, G. Margrave, D. Mitra, and K. R. Stanley, "Novel Techniques for Design and Control of Generalized Processor Sharing Schedulers for Multiple QoS Classes," in *IEEE Infocom*, 2000, pp. 932–941.

[12] J. C. Bennett and H. Zhang, "Why WFQ Is Not Good Enough for Integrated Services Networks," in *NOSSDAV*, April 1996, pp. 524–532.

[13] V. Sivaraman and F. M. Chiussi, "Providing End-to-End Statistical Delay Guarantees with Earliest Deadline First Scheduling and Per-Hop Traffic Shaping," in *IEEE Infocom*, 2000, pp. 631–640.

[14] L. Georgiadis, R. Guerin, and A. K. Parekh, "Optimal Multiplexing on a Single Link: Delay and Buffer Requirements," in *IEEE Infocom*, 1994, pp. 524–532.

[15] D. E. Wrege and J. Liebeherr, "A Near-Optimal Packet Scheduler for QoS Networks," in *IEEE Infocom*, 1997.

[16] M. Andrews, "Probabilistic End-to-End Delay Bounds for Earliest Deadline First Scheduling," in *IEEE Infocom*, 2000.

[17] Bavarian Archive for Speech Signals (BAS), "Verbmobil 6.1," http://www.phonetik.uni-muenchnen.de, 1996.

[18] 3GPP, "3G TS26.101 version 1.4.0: AMR Speech Codec Frame Structure - General Description," Oct. 1999.

[19] O. Rose, "A Memory Markov Chain Model For VBR Traffic With Strong Positive Correlations," in 16^{th} *ITC*, Edingburgh, United Kingdom, June 1999, pp. 827–836.