SERVICE DISCIPLINES PERFORMANCE FOR GPRS WITH LINK ADAPTATION AND HETEROGENEOUS TRAFFIC

Josep Bada and Fernando Casadevall

Universitat Politécnica de Catalunya , Jordi Girona 1-3, 08034 Barcelona- Spain, e-mail:ferranc@tsc.upc.es

Abstract - The General Packet Radio Service (GPRS) system provides packet switched data service with quality of service (QoS) over a GSM infrastructure. In order to cope with the QoS requirements, packet scheduling strategies are used. On the other hand, GPRS introduces four coding schemes with different degrees of data protection according to the radio channel conditions. Then, the selection of appropriate coding scheme according to the time variant channel conditions would lead to a substantial increment of the quality of service (QoS). In this paper, two scheduling policies, namely MED (Modified Earliest Deadline) and MLT (Minimum Laxity Threshold) are evaluated assuming both link adaptation (LA) strategy, based on BLER estimation, and heterogeneous traffic. Web browsing, fleet management, SMS (Short Message Service) and e-mail services are assumed. The performance of the proposed scheduling policies is evaluated in terms of throughput and delay. For urban environments with high load conditions, the obtained results show that a simple LA enhances QoS system performances, even though there is a trade-off with respect to the network packet loss.

Keywords – Mobile Communications, GPRS, Scheduling Strategies, Link Adaptation

I. INTRODUCTION

Nowadays, the success of wireless data is limited because certain requirements for data rates and cost are not properly fulfilled in a GSM network. However, it is not visionary to predict that with enhanced data services, e.g. GPRS (General Packet Radio Service), EDGE (Enhanced Data Rates for GSM Evolution) and UMTS (Universal Mobile Telecommunications System), this will change.

In the current mobile radio systems, data services are based on circuit switched radio transmission. For bursty traffic, this scheme is highly inefficient in terms of resource utilization. The bursty nature of data traffic makes packet switched technologies necessary if new and exciting wireless data services must be offered. For more than three years the ETSI (European Telecommunications Standards Institute) worked on the specification of GPRS. This system allows a service subscriber to send and receive data in an end-to-end packet transfer mode with effective peak data rates of some tens of Kbps.

For error correction purposes, GPRS defines four coding schemes (CS-1, CS-2, CS-3 and CS-4) in the RLC/MAC (Radio Link Control/Medium Access Control) layer. The performance of every coding scheme strongly depends on the channel quality conditions during the connection. To cope with the time variant propagation conditions, Link Adaptation strategies [1], which aim to maximize the channel throughput by using the most suited coding scheme at each time, could be considered. The throughput increment is important because it reduces the transmission delay and, therefore, it is possible to cope with the quality of service requirements for different data traffic applications, including short packets that have severe restrictions.

The link adaptation strategy assumed in this paper uses the Block Erasure Rate (BLER) as the channel quality indicator, [2]. The transmitter obtains the channel quality information from the RLC/MAC acknowledgement messages [3] that contain information of the successfully received radio blocks.

The system performances are analysed assuming urban conditions (a TU50 ideal FH radio channel is considered) and heterogeneous traffic. Two different scheduling (MED and MLT) disciplines, adapted to GPRS form ATM (Asynchronous Transfer Mode), are taken into account to fulfil the QoS requirements.

The paper is organized as follows. First, the overview of GPRS technology is presented. Section 3 describes the Link Adaptation strategy implemented. Section 4 describes in detail the scheduling algorithms analysed in the paper. The following section presents and discusses the obtained simulation results. Finally a set of conclusions is drawn.

II. GENERAL PACKET RADIO SERVICE

GPRS is an on-demand based bearer data service over the GSM air interface that offers quality of service to the users by dynamically allocation different amount of resource depending on the user requirements (throughput, delay, reliability, priority radio…). To introduce GPRS in the existing GSM infrastructure, additional network elements are added to the GSM architecture. This new structure is depicted in Figure 1. Basically, only two new nodes types, serving GPRS support node (SGSN) and gateway GPRS support node (GGSN), have to be introduced. In addition, this new technology requires the development of new mobile terminals.

The SGSN is responsible for the communication between the mobile station (MS) and the GPRS network. The GGSN provides the interface to external packet data network like X.25 or the Internet, but also to GPRS networks of other mobile operators. The GGSN node routes incoming packets to the appropriate SGSN for a particular MS. Within a Public Land mobile Network (PLMN) an IP-based backbone network is used for the communication between the GPRS nodes.

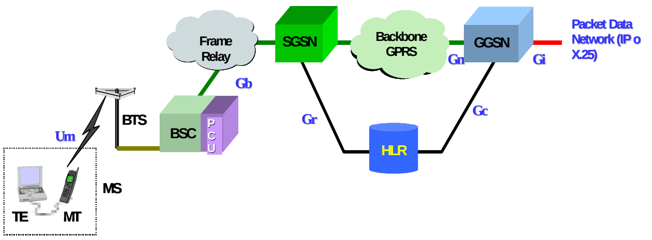

Fig 1. GPRS architecture

The GSM Base Station Subsystem (BSS) shares the radio resource between both circuit and packet switched services. Therefore, for packet data management through the radio channel the Base Station Controller (BSC) must also incorporate the so-called Packet Control Unit (PCU). The physical channels, available in a GSM cell, are dynamically shared between GPRS and GSM services. The ones associated with GPRS are called Packet Data Channel (PDCH). The basic transmission unit of a PDCH is called radio block. To transmit a radio block four time slots (TS´s), four consecutive TDMA frames are used.

A PDCH is structured in multiframes comprising of 52 TDMA frames, which corresponds to a duration of 240 ms. The mean transmission time per radio block is 20 ms. A radio block contains 456 bits (114 per burst). The structure and also the number of payload bits of a radio block depend on the message type and coding scheme [4] (Table 1).

Table 1. Coding parameters for the coding schemes

Channel Coding Scheme	Data bits in radio block	Data rate per time-slot (Kbps)
CS-1	181	9,05
CS-2	268	13,4
CS-3	312	15,6
CS-4	428	21,4

In order to use the scarce radio resource in a more efficient way and to support a number of applications with different requirements, GPRS provides several QoS profiles. These QoS profiles allow the mobile operators to create schemes for charging differentiation.

Table 2. The QoS Profile

Parameter	Values					
Precedence	Hight,normal, low					
Reliability	Packet loss probability, duplicate, out of sequence					
Delay	Size SDU	Class	1	2	3	4
	128 (bytes)	Mean(s)	<0,5	<5	<50	Best Effort
		95%(s)	<1,5	<25	<250	Best Effort
	1024 (bytes)	Mean(s)	<2	<15	<75	Best Effort
		95%(s)	<7	<75	<375	Best Effort
Rate	Mean	Depend				
	Maximum					

In Release'99, the QoS profiles are characterized by the five parameters listed in Table 2.

III. LINK ADAPTATION

The radio channel quality conditions may vary during the connection between the MS and the BSS. The goal of the link adaptation algorithm is the selection of the finest coding schemes according to channel conditions in order to reach a throughput as high as possible.

The proposed link adaptation algorithm is based on the use of the BLER parameter as the channel quality indicator. This parameter is evaluated at the end of every message of the window acknowledgement. The BLER estimation has been calculated using the following expression:

$$\overline{BLER(k)} = \overline{BLER(k-1)} \times (1-\alpha) + BLER(k) \times \alpha$$

where α =0.8 and k ε [1,window Ack size].

This expression is equivalent to a first order FIR filter and it is used to weigh the samples, being the weight of the most recent RLC/MAC data blocks higher. The α value controls the memory of the algorithm.

Once the BLER is estimated, the equivalent CIR is calculated according to current coding scheme. Figure 2 shows this relationship. Finally, the link adaptation algorithm moves to the finest coding scheme according to the Table 3 criteria. The new coding scheme is applied over the following LLC frames. Notice that this link adaptation algorithm is channel dependent.

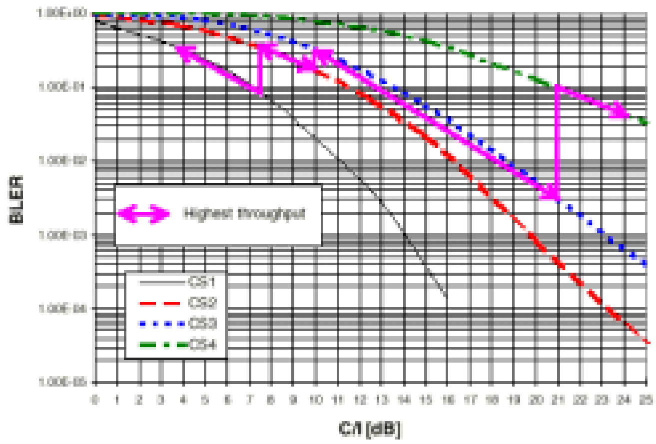

Fig 2. BLER vs C/I. Arrows indicate the highest throughput.

There are several different approaches that can be used to determine the initial coding scheme. In the downlink, the finest coding scheme is chosen in terms of the distance to the BTS. It can be proved that for the considered urban case, with a cell radio (Rc) of 3 Km and a frequency reuse parameter (K) equal to 4, the CS4 scheme is optimal for distances lower than 2010 meters while the CS3 scheme performs better for the interval 2010 meters < Rc ≤3000 meters.

The link adaptation is implemented only in downlink direction.

Table 3. C/I Thresholds

Finest Coding Scheme	CIR Threshold
CS1	CIR <= 7.5dB
CS2	7.5dB<CIR<=10dB
CS3	10dB<CIR<=21dB
CS4	CIR>21dB

IV. RADIO RESOURCE MANAGEMENT

In any wireless system, there is clearly a very scarce resource, which is the available bandwidth. In a GPRS multi-service environment, to guarantee the required QoS profiles and to optimise the system capacity, the use of efficient algorithms for Radio Resource Management (RRM), which will not be standardized because their specific realization is implementation dependent, becomes mandatory. The key element of these RRM algorithms is the scheduling algorithms, devoted to control the assignment of radio resources at the different queues for satisfying requirements of the all the active users. Our investigation is based on two comparative service disciplines with quality of service, namely: MLT (Minimum Laxity Threshold) [5] and MED (Modified Earliest Deadline) [6] and [7], that are used in ATM and have been properly adapted for the GPRS system .

A. MED

In the MED scheme, for each radio block belonging to packet to be scheduled, a transmission deadline is assigned. The radio blocks are served according to this deadline (d). The deadline value is function of quality of service parameters negotiated (delay). Denote by "t_p" the arrival time of a packet, the packet length by "l_p", its delay class by c, and the time-slot capability of the mobile terminal by "c_ms". Denote also the data rate of one time-slot, which depends on the used coding scheme, as "r_c". Then, the radio block deadline is computed by:

$$d = t_p + r(c, l_p) - \frac{l_p}{r\_c * c\_ms}$$

where, $r(c,l_p)$ is a function that depends on the size of the packet and the delay class "c", [8].

The MED strategy assumes four "late queues" (numbered from 0 to 3) and one "deadline queue" (Figure 3). The queues are served in the following order: first the "late queues", starting from 0 to 3, and finally the "deadline queue". The incoming radio blocks are placed in the "deadline queue" and arranged according to their deadlines values. If a radio block in the "deadline queue" is not transmitted before its deadline, then it is moved to the appropriate "late queues" according to its QoS (user priority).

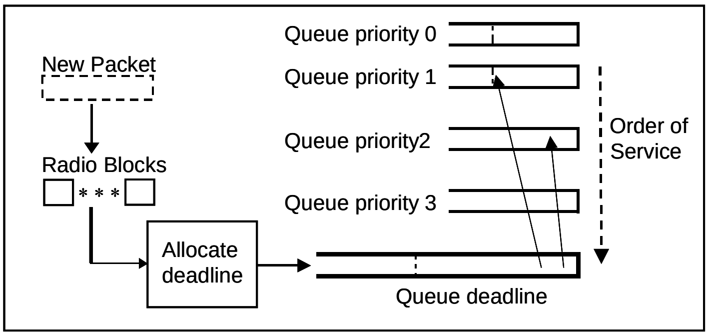

Fig 3. MED queuing scheme

B. MLT

The MTL policy introduces the laxity concept (Figure 4). The laxity represents the amount of time for which the PCU scheduler may remain idle, or serving radio blocks of other classes, and still it is able to transmit a given radio block before the end of its deadline. It is assumed that the radio blocks in the buffer are sorted according to their deadlines.

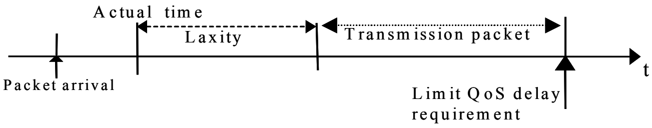

Fig. 4. Laxity concept

Denote as "RB_k" the number of radio blocks stored in the class k queue and "$d_k(j)$" the deadline of the j-th radio block of this class k. The laxity "$Lx_1(j)$" of the jth radio block allocated in the queue with the highest priority (Q1) at the time t is defined by:

$$Lx_1(j) = d_1(j) - t - [(j-1) \cdot tx(RB)]$$

where "tx(RB)" is the radio block time transmission (20ms).

The laxity for radio blocks allocated in a queue (Q2) with lower priority than the previous one is given by:

$$Lx_2(j) = d_2(j) - t - [RB_1 \cdot tx(RB)] - [(j-1) \cdot tx(RB)]$$

The above expression takes into account that there are "RB_1" radio blocks allocated in the queue Q1 that shall be previously served. A similar expression can be obtained for the Q3 priority queue.

Prior to each radio block transmission time, the laxities are evaluated for each of the radio blocks in the queues for traffic classes and the minimum laxities are computed for the queues themselves, that is: $L_K = \min_{0 \le i \le RB_i} (L_K(i))$. A threshold operation is the used to choose which QoS to serve. If $Lx_1 < 2*tx(RB)$ then queue high priority (Q1) must be served. If not, the minimum laxity of the second priority.

queue is compared. If this second priority queue cannot be served, then the queue with the third priority class is considered. If neither of these conditions is true, then best effort traffic may be transmitted

V. SIMULATION RESULTS

In order to examine the performance of the proposed scheduling and link adaptation algorithms, we have created a GPRS software simulator using MATLAB (Matrix Laboratory) simulation tool. To generate the different combinations of data traffic, four types of traffic sources are considered, namely: e-mail, WWW-WAP, fleet management and short message applications.

The model used for e-mail application is described in [9] and assumes that incoming messages are stored at a dedicated E-mail server. A comprehensive model for Web traffic, called the "behavioural model" was presented by Choi and Limb,[10]. In our study, we have adapted this model modifying the mean value of the IP packet length to emulate the comportment of WAP (Wireless Application Part) in the context of a GPRS system. The Fleet Management model is based on statistics obtained from fleet management applications using the Mobitex wireless packet data network in Sweden. The average message length on the uplink is 30 bytes and 115 bytes on the downlink [8]. Finally, the SMS, which is one of the most popular services in the actual GSM system, has been also included in our study. For SMS traffic, the length of the unit short message is limited to 140 bytes in order to take into account the restrictions in the Mobile Application Part (MAP) signalling layer.

For all the models, the packet or session inter-arrival time distribution is assumed to be exponentially distributed. In our analysis, it is assumed that there are 2 time slots (TS) exclusively dedicated for GPRS system in the cell. The maximum number of simultaneous active mobile stations in downlink direction is 32 and 15 in the uplink case. The mobiles terminals allocate (2:1) TS per connection. No IP header compression is used within the SNDCP. Service scheduling is performed at radio blocks level in the RLC/MAC, and it is assumed that this layer operates in acknowledge mode with a maximum number of retransmissions equal three. On the other hand, the LLC layer operates in non-acknowledge mode with maximum payload size of 1520 bytes. A TU50-noFH radio channel is assumed and the time simulation is one hour. All the MS are randomly distributed in the radio cell (R=3 Km). Table 4 depicts for each service the assumed specific QoS classes and the traffic percentage.

Figure 5 shows the normalized delay for QoS 1 class for both scheduling disciplines considering either Link Adaptation or fixed CS-1 coding scheme. It can be concluded that MED and MLT performs quit similar. When a fixed CS1 coding scheme is assumed, the link utilization can grow up to 70% still achieving the QoS delay requirements. However, when link adaptation is used, the link utilization increases substantially, being the maximum number of simultaneous MS the true system limitation. Notice that in the downlink direction the maximum number of simultaneous users is limited by TFI (Temporary Flow Identity), that is 32 active connections. The link utilization is defined as the degree of occupation of available resources for an hour, (i.e. if all resources are occupied for an hour then the link utilization is 1).

Table 4. Simulated scenarios

Services	Direction	Delay Class	Portion
WWW	Downlink	QoS 1	13%
E-mail	Downlink	QoS 1	23%
Mobitex	Downlink	QoS 2	39%
SMS	Downlink	Best Effort	26%
E-mail	Uplink	QoS 1	23%
Mobitex	Uplink	QoS 2	52%
SMS	Uplink	Best Effort	26%

Note.- In uplink direction only the Get Request message for has been assumed for the WWW service

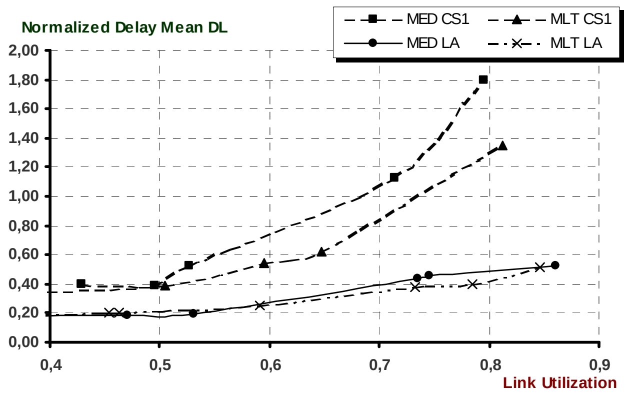

Fig. 5. Normalized delay vs. Link Utilization QoS 1

Figure 6 displays the normalized delay for QoS_2 and downlink direction. Since the Mobitex traffic source is very bursty with small packet length, when the link utilization increase in the case of fixed coding scheme, it is not possible to cope with the requested QoS delay, because the network falls into congestion (the normalized delay is higher than 1) for link utilization higher than 55% . However, when LA is used, the link utilization could increases up to around 65% for MLT and 75% for MED, without the networks falling into congestion state.

In the up link case, for medium and high link utilization values (higher than 50%), the normalized delay is higher than 1 (the system falls in congestion) for both QoS 1 and QoS 2 service classes (Figure 7). This is due because in the uplink case, multislot capability is limited to 1 TS and therefore the packets remain in the queue long time until it is transmitted.

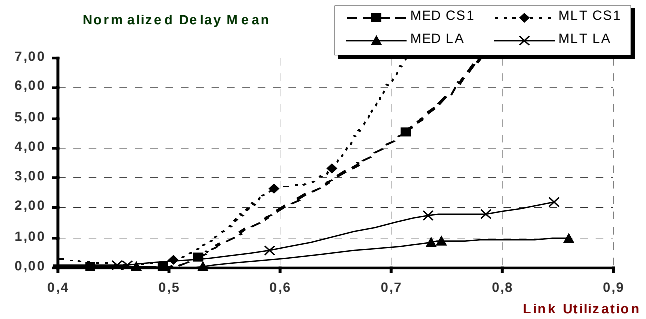

Fig. 6. Normalized delay vs. Link Utilization . QoS 2 service class

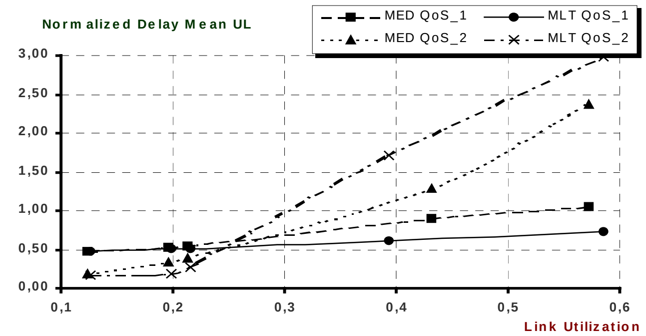

Fig. 7. Normalized Delay. Uplink case

Other QoS parameter studied in this paper is the maximum delay in the 95% of the cases. From Figure 8, it can be appreciated that this parameter less restrictive than the mean delay because, for all the services, the normalized delay is less than one for both fixed coding scheme (CS-1) and link adaptation strategies.

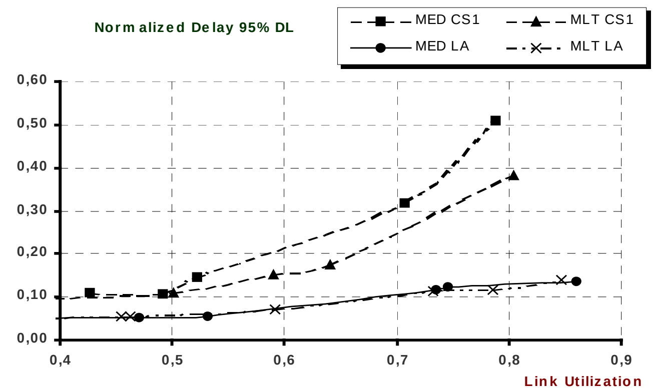

Fig. 8. Delay 95 vs. Link Utilization QoS 1

In order to have a complete vision on the behaviour of the proposed strategies, it is also important to analyse the mean waiting time (Delay 1 RB) before the transmission of the first radio block of a packet. From Figure 9, we can see that, for the MLT algorithm the mean waiting time is much lower than for the MED strategy when the traffic load increases. This can be explained as follows. For the case of high traffic load, the MED strategy serves the packets very near their deadline, even later when the network is in congestion. However, the MTL strategy takes into account the laxity of the packet. Packets with low laxity are sent faster. Similar results, as the shown in figure 9, are obtained when link adaptation is assumed.

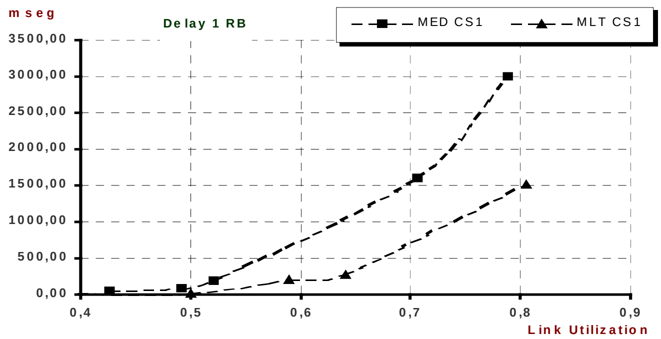

Fig. 9. Delay transmission first RB vs. Link Utilization

Figure 10 shows the evolution of the throughput as function of total number of sessions/hour (λ) for MED algorithm. The throughput is defined as the amount of error-free user data that reaches the destination during the simulation time. That is, the possible retransmissions at the RLC/MAC level are not taken into consideration. From the figure, it can be concluded that for e-mail applications is possible to increase the system throughput up to 110%, with respect to the use of a fixed CS1 coding scheme, when link adaptation is assumed, while for Mobitex traffic this increase is only of the 38%. These differences are due to packet length variation. For short packets, such as Mobitex traffic, the packet is allocated into two o eight radio blocks and, therefore, due to the reception of acknowledge messages the throughput increment is limited. In addition, when LA is used, as the coding scheme is adapted to the channel conditions, the throughput is practically constant (E-mail and Mobitex) because the delay bound is always below the congestion state.

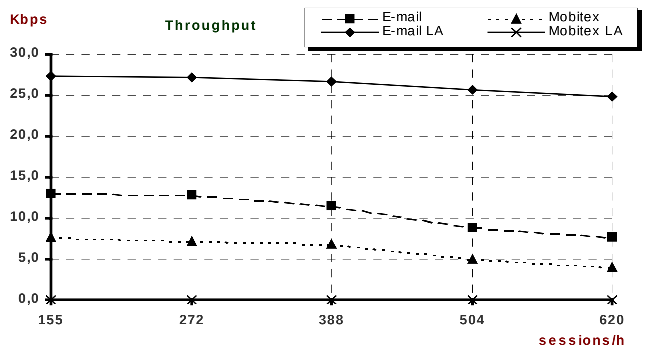

Fig. 10. Throughput vs. λ (sessions/h)

Figure 11 shows the evolution of the packet loss probability versus the total number of sessions/hour when link adaptation is used. It can be shown that packet loss is approximately constant around 4% with LA and very low (much less than 1%) for CS1 fixed coding scheme. The figure also shows the evolution of the number of radio block retransmitted (460 with CS-1, 27.000 with LA in the case of 620 sessions/hour). Notice that for LA, there is a trade-off between throughput increment and the loss probability.

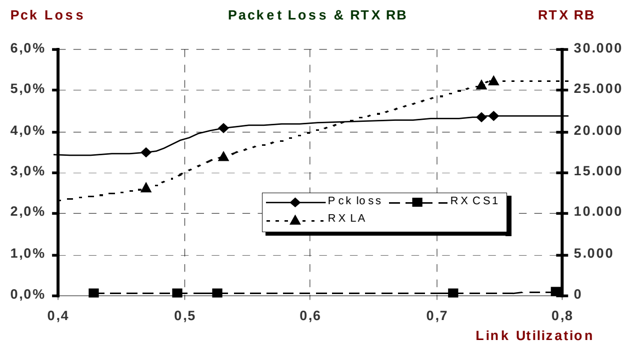

Fig. 11. Packet loss and RB retransmissions

Finally, for each pattern traffics and for both scheduling strategies, Table 5 shows the maximum number of the sessions/hour that can be admitted to cope with the requirements in terms of quality of service. We can realize that for LA, the parameters that limit the allowed number of sessions are the link utilization as well as the maximum number of simultaneous users.

Table 5. Comparative sessions/hour for MED and MLT

Discipline	**MED**		**MLT**	
Traffic	λ (CS1)	λ (LA)	λ (CS1)	λ (LA)
WWW	< 60	< 190	< 70	< 190
E-mail	< 105	< 330	< 150	< 330
SMS	See Note	See Note	See Note	See Note
Mobitex	< 165	< 570	< 140	< 570

VI. CONCLUSIONS

In this paper, we studied and evaluated the behaviour of different radio resource management methods for GPRS assuming heterogeneous traffic. The simulation results show that for the downlink case it is possible to achieve a high link utilization guaranteeing the delay requirements (QoS_1) with the proposed scheduling disciplines. MLT algorithm performs better than MED in terms of delay, despite of its higher implementation and processing complexity (10:1). Moreover, for urban conditions, when link adaptation is introduced, it has been shown that it is possible to increase the number of sessions/hour (Table 5) coping with the required delay bound. Moreover, in that case, a significant increment of the packet losses, up to 4%, happens. In order to solve this problem, a more accurate BLER estimation can be achieved if the length of LLC frame diminishes allowing the link adaptation FIR filter to perform.

ACKNOWLEDGEMENTS

This work has been carried out in the framework of the research project TIC-2001-2222, founded by CICYT-Spanish Science and Technology Ministry.

REFERENCES

[1] Ameigeiras P.J., Wigard J., Andersen P.N., Damgaard H.C., Mogensen P. "Performance of Link Adaptation in GPRS Network", VTC'2000, pp 492-499.

[2] Digital cellular telecommunications system (Phase 2+) (GSM);Background for Radio Frequency (RF) requirements, GSM 05.50 version 7.1.0 Release 1998

[3] Digital cellular telecommunications system (Phase 2+);General Packet Radio Service (GPRS); Radio Link Control/Medium Access Control (RLC/MAC) protocol, GSM 04.60 version 8.4.1 Release 1999.

[4] European Telecommunications Standards Institute, Digital cellular telecommunications system (Phase 2+) General Packet Radio Service- Overall description of the GPRS radio interface; Stage 2, GSM 03.64 version 7.0.0 Release 1998

[5] Jay M. Hyman, Aurel A. Lazar and Giovanni Pacific: "Real-Time Scheduling with Quality of Service Constraints", IEEE Journal on Selected Areas in Communications, September 1991, Vol. 9, No 7, pp. 1052-1063

[6] Duke Hong and Tatsuya Suda, "Congestion Control and Prevention in ATM Network", IEEE Network Magazine July 1991, pp 10-16

[7] Hiroshi Saito, "Optimal Queuing Discipline for Real-Time Traffic at ATM Switching Nodes", IEEE Transactions on Communications, December 1990, Vol. 38, No 12, pp. 2131- 2136

[8] Pang Q., Bigloo A., Leung V. Sccholefield C. ., "Service Scheduling for General Packet Radio Service", IEEE Wireless Communications and Networking Conference 1999, September 1999, New Orleans, pp. 1229-1232

[9] Ho Joseph, Zhu Yixin and Madhavapeddy Seshu "Throughput and Buffer Analysis for GSM General Packet Radio Service (GPRS)" IEEE Wireless Communications and Networking Conference 1999, September 1999, New Orleans, pp. 1427- 1431

[10] Choi, H., Limb, J. A behavioural model of web traffic. ICNP 99, Sep 1999.