Novel Scheduling Algorithms Aiming for QoS Guarantees for VoIP over HSDPA

André R. Braga , Emanuel B. Rodrigues and Francisco R. P. Cavalcanti

Abstract— **The growth in the number of Voice over IP (VoIP) users on the internet makes it the service with the highest interest to be provided by cellular operators. On the other hand, it demands very strict QoS control, which becomes even more complicated in wireless networks, because packets can be lost due to radio link transmission errors, as well as network congestion. Within this paradigm, scheduling strategies appear as the most suitable solution to cope with QoS guarantees under high loads, where the resources should be carefully distributed among the users in order to fulfill their QoS requirements.**

This work comprises the proposal and evaluation of some scheduling algorithms aiming to improve system capacity and QoS guarantees for speech users in HSDPA.

The results show that the scheduling algorithms that mix fairness with channel information to get improvements of the multi-user diversity provide the best performance.

Index Terms— **VoIP, HSDPA, QoS, all-IP, scheduling.**

I. Introduction

In traditional wireless networks, Real-Time (RT) services (e.g., voice) are carried over dedicated channels because of their delay sensitivity while Non-Real-Time (NRT) services (e.g., web browsing) are transported over time-shared channels because of their burstiness. It has recently been proposed that even RT services can be efficiently transported over time-shared channels supported on the forward link of UMTS.

Some works have dealt with Packet Scheduling (PS) algorithms in HSDPA. The performance of NRT services (web browsing, FTP) was evaluated in [1], for example. The evaluation of several PS techniques for a mixed traffic scenario composed of RT service (streaming) and NRT service (web browsing) was presented in [2], [3], [4]. The main focus of the present study is to evaluate the performance of the VoIP service. Previous works have studied the performance of this specific service alone, in a single service scenario [5], [6]. The only work that has assessed the VoIP service was [7], but its focus was the comparison between voice on DCH and High Speed Downlink Share Channel (HS-DSCH), and not the complete evaluation of several PS algorithms. The novelty of the present work comes from the fact that it performs a complete evaluation of the VoIP service for several PS algorithms in terms of QoS and capacity.

The results presented in [8], [9] provide evaluations of scheduling algorithms in mixed traffic scenarios and different end-to-end delay requirements as well.

The authors are with the Wireless Telecommunications Research Group (GTEL), Federal University of Ceará (UFC), CP 6005 - Campus do Pici - 60455-760, Fortaleza-CE , Brazil. E-mails: { andrerb, emanuel, rodrigo}@gtel.ufc.br.

The remainder of this paper is organized as follows. Section II presents the mathematical formulation of all the evaluated PS algorithms throughout this study. The simulation results as well as the simulation modelings are depicted in Sections III and IV, respectively. Finally, some discussions, conclusions are drawn in Section V.

II. Scheduling Algorithms

The scheduling problem in wireless systems has a basic characteristic that distinguishes it from wireline systems: the link variability. That is the reason why this technique plays an important role in providing packet switched data services.

The algorithms employed should be aware of several issues like the radio link, to avoid wasting resources in transmissions during error states; fairness, trying to guarantee a certain service rate to a flow [10]; QoS by means of differentiated services that should be given different priorities; and also the complexity, that must be low enough to be feasible to be implemented in real communication systems.

This section describes each algorithm employed in this study. Two sets of algorithms were used: QoS-differentiated and non-QoS-differentiated algorithms. The latter cannot differentiate between services or QoS demands of each specific user. The capabilities of all the presented schedulers in providing QoS to both VoIP and Web services are evaluated later in this chapter.

The description of each algorithm comprises the calculation of the priority function for each user. One should keep in mind that, for all the algorithms, retransmissions are given higher priority over scheduling of new data. This priority is expressed by means of a multiplicative factor applied to the calculated priority.

A. Non-QoS-differentiated algorithms

1) Round Robin (RR): The users are served in a cyclic order ignoring the channel conditions [4]. The priority calculation is based on the queueing time of each users. It is important to mention that the queueing time of each user is only updated if the transmission buffer is not empty. This algorithm provides a fair resource distribution among all users in the queue.

2) Proportional Fair (PF): This algorithm intends to serve users with favorable radio conditions in providing a high instantaneous throughput relative to their average throughput [4]. The priority value for user i is calculated based on Equation (1):

$$p_i = \frac{R_i}{T_i}\ , \tag{1}$$

where p_i is the priority value for user i, R_i is the estimated bit rate for the next transmission attempt and T_i is the average supported throughput.

3) Queue Based Max CIR (QBMC): This algorithm incorporates to the classic Max CIR algorithm, information regarding the queue size of each users transmission buffer. Some other algorithms proposed in [4] and [11] took into account the queue size information as well. With this information, the scheduler is able to prioritize queues that are not being served due to their channel conditions. In fact the queue length is an indirect measure of the delay. The priority calculation of user i is presented in Equation (2):

$$p_i = C_i.B_i \,, \tag{2}$$

where C_i is the measured CIR and B_i is the buffer size in number of bits not yet transmitted.

B. QoS-differentiated algorithms

1) Linear and Asymptotic Delay Scheduler (LDS and ADS): The basic concept of the ADS was proposed by [12] using a Barrier Function (BF) which value is multiplied by the PF function. This results in a priority function which is aware of the delay requirements of each specific user as well as of the ratio between instantaneous and average bit rate. It is specially suited for VoIP. Equation (3) presents the priority calculation for each user i:

$$p_i = \frac{R_i}{T_i}.B_i \,, \tag{3}$$

where B_i is the BF for user i, which can be calculated by using Equations (4) or (5). They represent an asymptotic and linear BF calculation, respectively:

$$B_i^{ADS} = 1 + \frac{1}{D_{th} - D_i} \,, \tag{4}$$

$$B_i^{LDS} = \frac{99}{D_{th}}.D_i + 1 \,, \tag{5}$$

considering D_{th} as the delay threshold and D_i is the head-of-line packet delay for user i. When D_i is equal to or larger than D_{th}, a constant value of 100 is assumed for the BF.

The LDS is also based on some other works [4] that use a linear component of the delay to be incorporated into the priority function.

The values assumed for D_{th} are based on each service class requirement. The configuration of this parameter makes it an algorithm that matches the delay requirement of each specific service. The BF calculation presented above has an advantage of, when a service does not have delay requirements, D_{th} can be set to a big value, so that the BF has a value close to 1, leading the priority calculation for the service to be equal to the PF.

Figure 1 presents the behavior of both BF considering $D_{th} = 150\ ms$. It can be noticed that the linear function presents higher values for low delay while the asymptotic function really increases only when the delay is close to the threshold value.

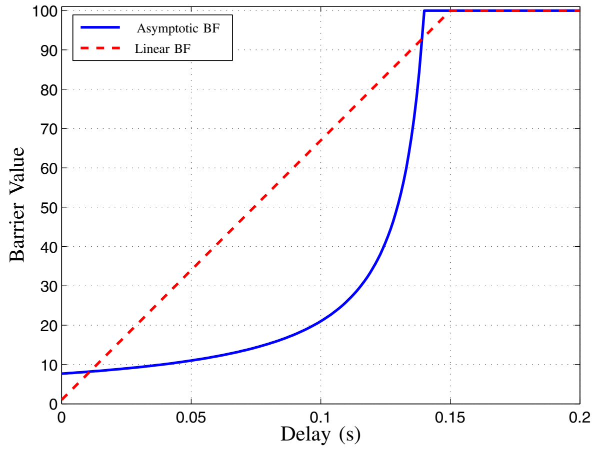

Fig. 1. BF behavior with increasing delay.

TABLE I

MAIN SIMULATION PARAMETERS.

Parameter	Value	Unit
Parallel HARQ Processes	6	-
Maximum HARQ RTx	5	-
HARQ type	Chase Combining	-
Max. # of Code Mult.	4	-
# Codes for HSDPA	5	-
Max. BS Tx Power	20	W
Common Channels Power	1	W
CPICH Power	2	W
HS-SCCH Power	0.5	W
AMR Codec Rate	12.2	kbps
Mean Holding Time	60	s
Compressed Header Size	48	bits
VoIP Discard Threshold	140	ms

In this work, the delay threshold assumed by the scheduler is the same as the adopted by the discard mechanism at the MAC-hs [13]. For VoIP services, the value depends on the delay budget configuration. When web users are considered, the threshold is considered to be close to infinite.

III. SIMULATION MODELLING

The scheduling algorithms are evaluated by means of a WCDMA radio network simulator featuring the HS-DSCH and built using C++. Table I presents the main simulation parameters assumed in the performed simulations.

It should be emphasized that the values of the capacity and quality figures depicted in the following graphics should not be considered as absolute performance indicators. The reader should focus on the presented relative comparisons, since the results depend highly on the specific parameters defining the scenario and are generally only illustrative. Performance of real products and systems may differ considerably.

The coverage area is composed of macrocell sites, where each site is formed by three hexagonal sectors/cells. A site-to-site distance of 1.5 km is also considered, i.e. cell radius is equal to 500 m.

The system comprises 27 cells, which all generate statistics throughout the simulation. In order to avoid border effects in the interference calculation, a wrap around technique [14] is used. Thus, the obtained simulation results are reliable in a statistic way and the model can represent a practical cellular system.

For the macrocell environment, the average propagation loss follows the assumptions for the Vehicular Test Environment, which is one of the UMTS operating environments described in [15]. This deployment model is characterized by larger cells and high transmit power. All the simulation scenarios are constructed utilizing the UMTS description and characterization found in [15].

A. Traffic model

The studied scenario is comprised exclusively of VoIP users. Within this scenario, the performance of the scheduling algorithms is analyzed focusing on their capabilities of sharing the resources among VoIP users.

The VoIP traffic pattern is based on a Two-State Voice Traffic Model. The 3GPP has chosen the AMR codec to be used in the VoIP service [16]. During ON periods, with AMR mode 12.2 kbps, the VoIP application generates 32-byte voice payload at 20 ms intervals.

In general, VoIP uses three layers of protocols, namely RTP [17], UDP [18] and IP [19], creating a massive header overhead of 40 bytes (for the case of IPv4). Such amount of overhead compared to the voice payload is highly unacceptable in a wireless environment. To solve this problem, header compression techniques are applied, for example ROHC, which can reduce the RTP/UDP/IP packet headers from 40 bytes to 2 bytes most of the time when UDP checksums are disabled, and to 4 bytes when UDP checksums are enabled [20].

According to [6], in order to achieve an acceptable quality for the VoIP call, the one-way mouth-to-ear delay should be less than 250-300 ms. This total delay should account for all the nodes in the communication path: the voice codec, local access network or PSTN, Internet, GGSN, SGSN, RNC, Node B and UE. The present research, is interested at the delay budget inside UTRAN (RNC, Node-B and UE). This delay budget should be enough for all the Node B functionalities (scheduling, HARQ procedures, etc.) and the user reception of VoIP packets.

Studies conducted in [6] considered delay budgets in the range from 80 ms to 150 ms, that can be supported by the RNA. This range should be sufficient for scenarios where the VoIP call is between two mobiles or between a land-line and a mobile user. Therefore, three delay budgets are considered in the simulations: 150, 130 and 110 ms.

B. Performance metrics

A VoIP user is assumed as satisfied if it has a Frame Erasure Rate (FER) lower or equal to 1%, reflecting a good perceived speech quality provided by the AMR codec with 2% FER (1% guaranteed for each link direction) [21].

The FER metric is concerned with the VoIP service only. A packet loss can occur in three different ways:

- Packet loss due to channel errors: this type of loss occurs during normal transmissions using the unacknowledged (UM) RLC mode. HSDPA includes HARQ at Node B level, but is limited to 5 HARQ attempts in this study. If the transmission quality is so bad that even several HARQ retransmissions are not enough to enable a correctly received packet, these transmission errors will ultimately result in packet losses. A secondary performance metric is calculated specifically for this kind of loss: Lost packets ratio due to channel errors;
- Packet loss due to discard at the MAC-hs layer: the MAC-hs layer employs a discard mechanism for the MAC-hs Service Data Unit (SDU). A packet is discarded when it arrives at the scheduler too late, i.e., its delay is higher than a discard timer. Since the packets are discarded before the actual transmission over the air interface, the resources required to deliver the packet are released and can be used for the remaining packets. A secondary performance metric is calculated specifically for this kind of loss: Discarded packets ratio at the MAC-hs layer;
- Packet loss due to discard at the play-out buffer: if the packet is correctly received but its delay is higher than the maximum allowed end-to-end delay (delay budget), the packet is also referred to as lost. This type of loss can be seen as if the terminal play-out buffer size is too small to handle the packets with excessively high delays and, therefore, these packets are thrown away. Note that these packets are not dropped until after the reception and will therefore consume resources. A secondary performance metric is calculated specifically for this kind of loss: Discarded packets ratio at the play-out buffer.

It is relevant to say that the FER metric calculation takes into account all these three kinds of packet loss. It is the ratio of the number of lost packets (voice frames) due to any of the three kinds of loss described above, and the total number of generated packets (voice frames).

IV. Performance Results

Firstly, the algorithms are evaluated by means of the satisfaction ratio and considering three different delay budgets at the application layer: 150, 130 and 110 ms. It is important to mention that the discard threshold at the MAC-hs is also tuned in each of these configurations. It is assumed to be always 10 ms less than the delay budget, i.e. 140, 120 and 100 ms, respectively. This is due to the fact that, when a MAC-hs SDU is concatenated into a MAC-hs Protocol Data Unit (PDU), it cannot be discarded anymore [6]. This gap difference tries to give 10 ms for the system to accomplish the transmission of the packet before the delay limit at the application.

In Figure 2 the percentage of satisfied users is presented and the scheduling algorithms can be compared with a delay budget of 150 ms. It can be noticed that the PF scheduler provides the worst performance. Since it only uses rate information to sort users in priority order, it cannot provide QoS guarantees for VoIP. The RR provides much better results, due to its property of sharing equally the transmission slots (TTIs) among all users in the queue.

The non-classical algorithms evaluated provide similar performance among them, and better results than the RR, mainly under loads beyond the capacity limit. They are more able to provide QoS guarantees under overloaded situations than classical algorithms.

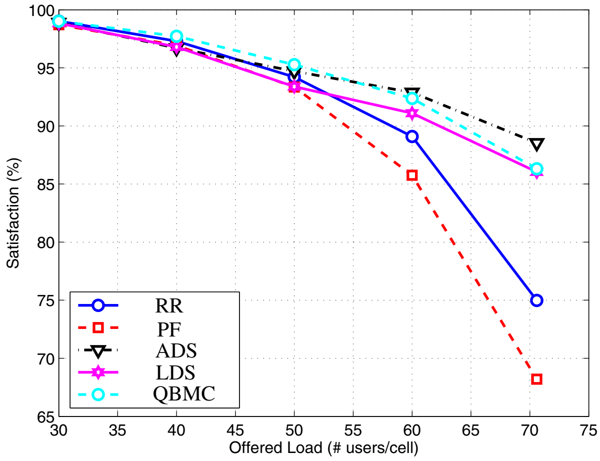

Fig. 2. VoIP satisfaction ratio with a 150 ms delay budget.

The 130 ms delay budget is evaluated in Figure 3. Comparing to both ADS and QBMC, the LDS provides the worst performance. Comparing to the case where the delay budget is set to 150 ms, the performance difference of these last two algorithms is smaller. It is also easy to perceive that the relative behavior between both the RR and PF has not changed significantly.

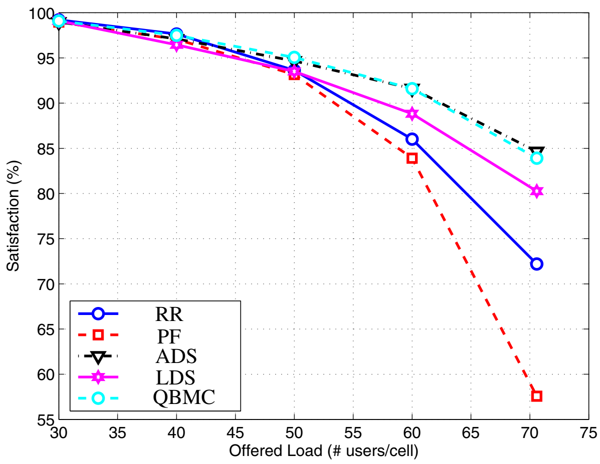

Fig. 3. VoIP satisfaction ratio with a 130 ms delay budget.

But when the delay budget is even smaller, as shown in Figure 4, the QBMC becomes the best algorithm for the highest simulated load. It is more evident that it provides a better trade off between channel usage, by the Max CIR component, and fairness by the queue size multiplicative factor. It means that, for this scenario, the QBMC provides the better relative performance under overloaded situations when the VoIP service to be provided has a lower delay requirement.

Another reason why both ADS and LDS suffers a performance degradation with a smaller delay budget is the fact that the system operation point becomes closer to the delay threshold. It leads to a higher influence of the BF in the priority calculation. Hence, the channel information becomes less important and, the scheduling process becomes less efficient from a channel usage point-of-view.

Comparing Figures 2 and 4 for the highest simulated load, the LDS has just about 5% of increase in satisfaction compared to the RR algorithm assuming a 110 ms delay budget, while more than 10% is obtained considering a higher delay budget. Since the system operates more frequently with packet delays closer to the threshold value, this scheduler forces more often a user to be scheduled under not ideal channel conditions.

As expected, the overall system performance has suffered a degradation with the decrease of the delay budget.

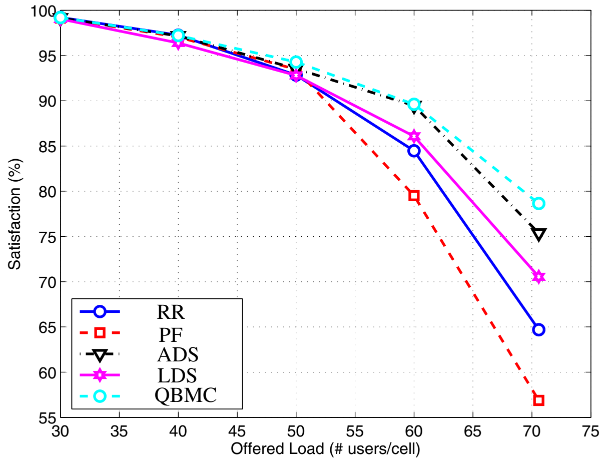

Fig. 4. VoIP satisfaction ratio with a 110 ms delay budget.

Some performance metrics can also be evaluated aiming at a detailed explanation of the characteristics regarding each algorithm. The 98^{th} percentile of lost packets ratio due to channel errors for each delay budget is presented in Figure 5. Since a packet is lost in HSDPA when the maximum number of transmission attempts is reached without successful reception, it shows how each algorithm is able to give access to the transmission, users experiencing the better channel conditions. The figure shows that all the algorithms provide ratios approximately between 1.0% and 2.5% for the highest simulated load.

A behavior that seems to be strange is that, the 98^{th} percentile of lost packets ratio of the RR, PF and QBMC algorithms does no increase comparing the highest two loads. It can be easily explained by the high discard ratios at the MAC-hs presented by these algorithms, which is presented in Figures 7 and 8. Since many packets are discarded at the transmitted, less packets are left to be lost at the channel.

Sometimes the 98^{th} percentile metric may not be clear about the performance regarding the desired metric. That is why in Figure 6 the CDF of the lost packets ratio regarding the highest simulated load is shown aiming at a more reliable conclusion to be reached.

Although the RR provides the better 98^{th} percentile performance, the CDF shows something different. Focusing into the 1% of lost packets ratio area, the best algorithms are the PF and QBMC. The RR has a closer performance compared to both the ADS and LDS. It indicates that the former two algorithms are more able to give access to transmission users experiencing better channel conditions, leading to a lower number of lost packets due to inefficient reception. The RR does not look at the channel conditions and both the ADS and LDS give a higher importance to the delay attribute during overloaded conditions, becoming the channel state less relevant for scheduling purposes.

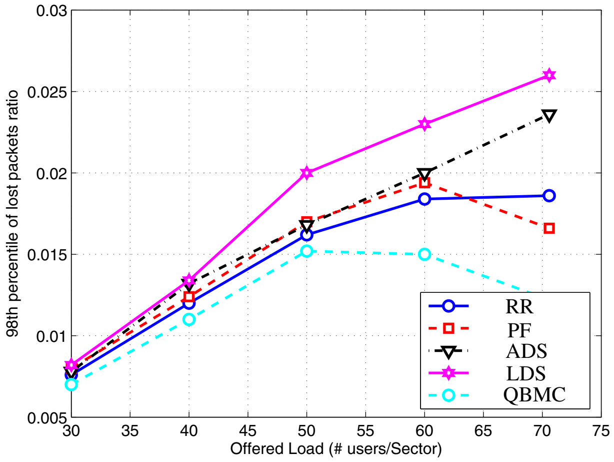

Fig. 5. Lost packets ratio with a 150 ms delay budget.

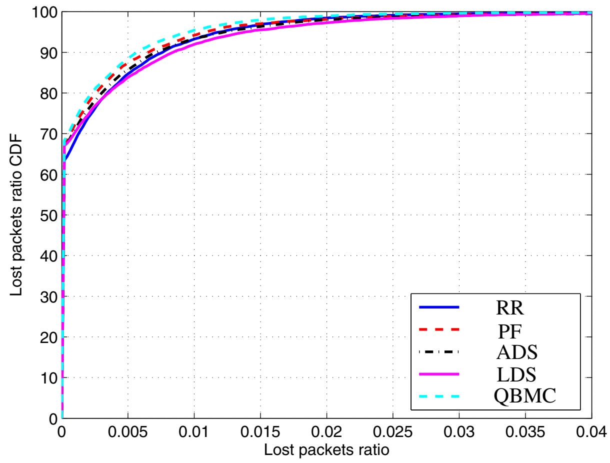

Fig. 6. CDF of the lost packets ratio for the highest load with a 150 ms delay budget.

The discard ratio at the MAC-hs is analyzed in Figure 7 through the 98^{th} percentile. It can be noticed that both the ADS and LDS provide the better performance. The former provides the smallest discard ratio. The RR is outperformed by the PF only under the highest evaluated load. Indicating higher queueing times at the HARQ buffers occasioned by the inexistance of channel knowledge at the scheduling process, leading to high retransmission rates.

The CDF of the discard ratio at the MAC-hs is depicted in Figure 8. It can be seen that the RR and PF provide the worst performance. Although the QBMC is also not aware of the QoS requirement for VoIP, its performance is quite similar to both the ADS and LDS, which monitor the delay metric and requirement of this service.

Now, the discard ratio at the play-out buffer in the application layer is evaluated in Figure 9. The RR provides always the worst performance with much higher discard ratios than the

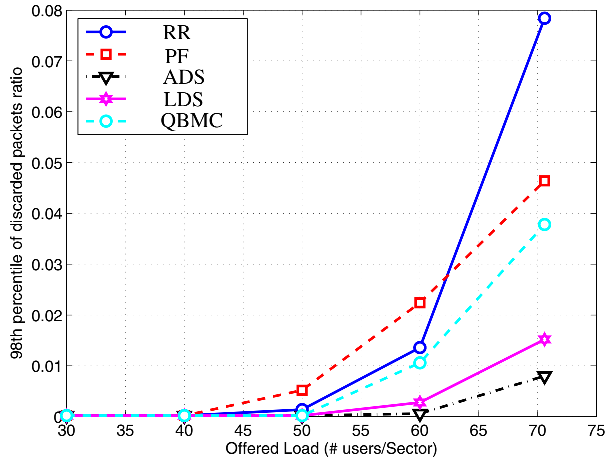

Fig. 7. Discarded packets ratio at MAC-hs with 150 ms delay budget.

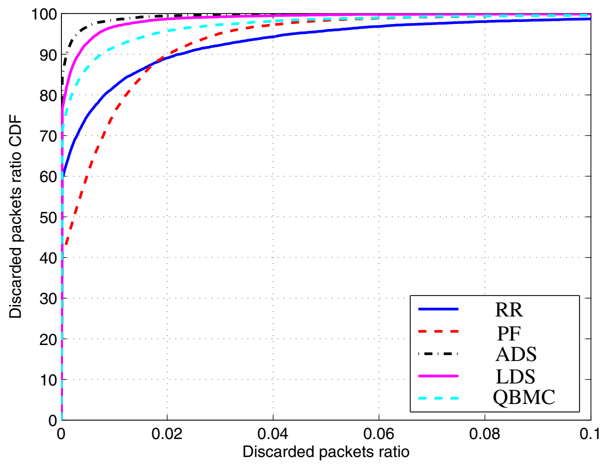

Fig. 8. CDF of the discarded packets ratio at MAC-hs for the highest load with 150 ms delay budget.

other algorithms. The PF shows itself as the best algorithm, regarding this metric. This is mainly due to the fact that it only schedules users with reasonable good channel conditions, leading to a lower number of required retransmissions to accomplish the correct transmission of a packet.

It is worth to mention that the scheduling algorithm is able to control the discard ratio at the MAC-hs, since it is able to force a user to be scheduled, avoiding it to have packets discarded. On the other hand, if the user is forced to be scheduled during bad channel conditions, it may take too much time to transmit the packet due to a big number of required retransmissions, or even expire the maximum number of transmission attempts, leading to a lost in the packet. The higher amount of retransmissions may also lead to a higher discard ratio at the application layer (play-out buffer).

The CDF of the discard ratio at the play-out buffer is shown in Figures 10. The CDF shows that the RR performance is the worst, as expected from an algorithms that is not aware of the channel conditions, resulting in a higher probability of longer times for the reception accomplishment of a packet due to a higher number of retransmissions during bad channel conditions. The other algorithms provide very similar performance.

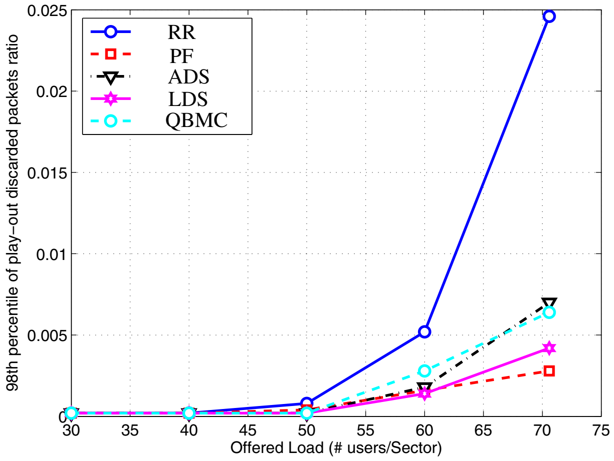

Fig. 9. Discarded packets ratio at the application with 150 ms delay budget.

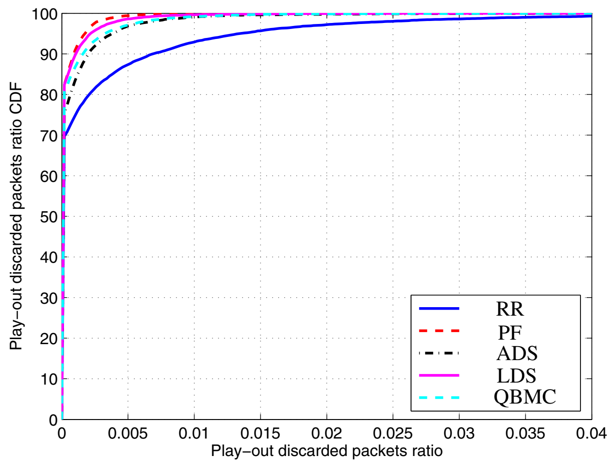

Fig. 10. CDF of the discarded packets ratio at the application for the highest load with 150 ms delay budget.

V. Conclusions

The main contribution of this work is the proposal and evaluation of the capability of some scheduling algorithms in providing QoS to the VoIP service offered in HSDPA.

It can be concluded that, although the PF algorithm provides very bad results due to its unawareness of the delay, the algorithms that obtained the better results are some sort of modified PF. The algorithms that used a barrier function provide the better overall results. In fact, they provide huge capacity gains compared to classical algorithms like the PF and RR.

The QBMC, which is a modified Max CIR algorithm, provides a remarkable performance as well. It has even an advantage when the delay requirement becomes stricter, turning out to be the best algorithm for overloaded conditions.

An interesting perspective is the evaluation of these algorithms considering a mixed traffic scenario, where the resources are shared among services with different requirements.

Acknowledgement

This work was supported by a grant from Ericsson of Brazil - research branch under ERBB/UFC.17 technical cooperation contract. The authors would like to thank the valuable comments of Sven-Olof Jonsson, Stefan Wänstedt and Mårten Ericson from Ericsson Research in Luleå, Sweden.

Francisco R. P. Cavalcanti was partly funded by CNPq - Conselho Nacional de Desenvolvimento Cientfico e Tecnolgico, grant no. 304477/2002-8.

References

[1] M. Kazmi and N. Wiberg, "Scheduling Algorithms for HS-DSCH in a WCDMA mixed traffic scenario," *14^{th} IEEE International Symposium on Personal, Indoor and Mobile Radio Communication - PIMRC*, pp. 1485–1489, 2003.

[2] M. Lundevall, B. Olin, J. Olsson, N. Wiberg, S. Wänstedt, J. Eriksson, and F. Eng, "Streaming Applications over HSDPA in Mixed Services Scenarios," *60^{th} IEEE Vehicular Technology Conference - VTC Fall*, September 2004.

[3] J. Eriksson, "Providing Quality of Service for Streaming Applications in Evolved 3G Networks," Master's thesis, Linköping University, Linköping, Sweden, January 2004.

[4] P. J. A. Gutiérrez, "Packet Scheduling and Quality of Service in HSDPA," Ph.D. dissertation, Aalborg University, October 2003.

[5] S. Wänstedt, F. Rui, M. Ericson, and M. Nordberg, "Providing Reliable and Efficient VoIP over Cellular Networks," in *Future Telecommunications Conference*, October 2005.

[6] B. Wang, K. I. Pedersen, T. E. Kolding, and P. E. Mogesen, "Performance of VoIP on HSDPA," in *IEEE Vehicular Technology Conference, Spring*, May 2005.

[7] N. Lithammer, "Speech on HS-DSCH," Master's thesis, Royal Institute of Technology, Sweden, Stockholm, December 2003.

[8] A. R. Braga, E. B. Rodrigues, and F. R. P. Cavalcanti, "Packet Scheduling for VoIP over HSDPA in Mixed Traffic Scenarios," *17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications - PIMRC*, September 2006.

[9] ——, "Packet Scheduling for Voice over IP over HSDPA in Mixed Traffic Scenarios with Different End-to-End Delay Budgets," *VI International Telecommunications Symposium - ITS*, September 2006.

[10] Y. Cao and V. O. K. Li, "Scheduling Algorithms in Broad-Band Wireless Networks," *Proceedings of the IEEE*, pp. 76–87, January 2001.

[11] P. Falconio and P. Dini, "Design and Performance Evaluation of Packet Scheduling Algorithms for Video Traffic in the High Speed Downlink Packet Access," in *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2004. PIMRC 2004*, vol. 3, September 2004, pp. 2202–2206.

[12] P. Hosein, "Scheduling of VoIP Traffic over a Time-shared Wireless Packet Data Channel," in *IEEE International Conference on Personal Wireless Communications (ICPWC)*, January 2005, pp. 38–41.

[13] 3GPP, "Medium Access Control (MAC) Protocol Specification," 3^{rd} Generation Partnership Project, Sophia Antipolis, France, Tech. Rep. TS 25.321 v6.7.0 - Release 6, December 2005. [Online]. Available: http://www.3gpp.org

[14] J. Zander, S.-L. Kim, M. Almgren, and O. Queseth, *Radio Resource Management for Wireless Networks*. Artech House Publishers, April 2001.

[15] UMTS, "Selection Procedures for the Choice of Radio Transmission Technologies of the UMTS," ETSI, UMTS TR 101.112 v.3.2.0, Tech. Rep., April 1998.

[16] 3GPP, "Packet Switched Conversational Multimedia Applications: Default Codecs," 3^{rd} Generation Partnership Project, Sophia Antipolis, France, Tech. Rep. TS 26.235 v7.0.0 - Release 7, December 2005. [Online]. Available: http://www.3gpp.org

[17] IETF, "RTP: A transport protocol for real-time applications," The Internet Engineering Task Force, Tech. Rep. RFC 3550, 2003.

[18] ——, "User datagram protocol," The Internet Engineering Task Force, Tech. Rep. RFC 768, 1980.

[19] ——, "Internet protocol," The Internet Engineering Task Force, Tech. Rep. RFC 791, 1981.

[20] ——, "Robust Header Compression (ROHC)," The Internet Engineering Task Force, Tech. Rep. RFC 3095, 2001.

[21] 3GPP, "Performance Characterization of the Adaptive Multi-Rate (AMR) Speech Codec," 3^{rd} Generation Partnership Project, Sophia Antipolis, France, Tech. Rep. TR 26.975 v6.0.0 - Release 6, December 2004. [Online]. Available: http://www.3gpp.org