Packet Scheduling for Voice over IP over HSDPA in Mixed Traffic Scenarios with Different End-to-End Delay Budgets

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

Abstract— **By means of dynamic system-level simulations, this work evaluates whether the studied Packet Scheduling (PS) algorithms are capable of guaranteeing the Quality of Service (QoS) of the Voice over IP (VoIP) on a scenario where the VoIP and web browsing services will compete for the same resources and there is a maximum allowed end-to-end delay for VoIP packets. Two sets of algorithms were considered: QoS-differentiated and non-QoS-differentiated algorithms. It was verified the system capacity gains when the VoIP delay budget is increased. Regarding the PS performance, those algorithms able to perform QoS differentiation presented better performance in terms of system capacity and QoS compared to those non-QoS-differentiated algorithms. Among the former category, those that take into consideration delay requirements of the VoIP service presented the best overall capacity for all traffic mixes and delay budgets.**

Index Terms— **High Speed Downlink Packet Access, Packet scheduling, Voice over IP, delay budget.**

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 High Speed Downlink Packet Access (HSDPA). A potential advantage of transmitting speech on a channel previously designed for data traffic is the improved efficiency in terms of resource sharing, spectrum usage, provision of multimedia services and network architecture. However, the challenge is to port VoIP services on wireless networks while retaining the QoS of today's circuit-switched networks and the inherent flexibility of IP-based services.

PS algorithms that support QoS differentiation and guarantees for wireless data networks are crucial to the development of broad-band wireless networks. The design of scheduling for wireless networks with QoS guarantees for the majority of users is particularly challenging, mainly because of limited resources and fast changing radio environments.

The performance of the VoIP service is very dependent on the considered delay budget, which accounts for the maximum allowed end-to-end (one-way mouth-to-ear) delay. This total delay should account for all the nodes in the communication path. The present research, is interested in the delay budget inside UMTS Terrestrial Radio Access Network (UTRAN) (Radio Network Controller (RNC), Node-B and User Equipment (UE)). This delay budget should be enough for all the Node B functionalities (scheduling, Hybrid Automatic Repeat Request (HARQ) procedures, etc.) and the user reception of VoIP packets. The delay budget is associated with a trade-off between packet loss due to excessive delay and interactivity between the communicating parties, which has a direct impact on system capacity and QoS.

Some works have dealt with PS algorithms for VoIP service on HSDPA in a single service scenario with different delay budgets [1], [2], [3], [4], [5]. Few works have assessed the VoIP service in a mixed traffic scenario [6], [7]. Paper [6] has evaluated only one delay budget, while paper [7] has focused at the comparison between voice on Dedicated Channel (DCH) and High Speed Downlink Shared Channel (HS-DSCH), and not the complete evaluation of several PS algorithms and different VoIP delay budgets. 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, considering mixed traffic scenarios and different end-to-end VoIP delay budgets. By means of dynamic system-level simulations, this work evaluates whether the studied PS algorithms are capable of guaranteeing the QoS of the VoIP on a scenario where the VoIP and web browsing services will compete for the same resources and there is a maximum allowed end-to-end delay for VoIP packets.

The remainder of this paper is organized as follows. Section II presents the mathematical formulation of all the evaluated PS algorithms. The description of the models implemented in the simulation tool is presented in Section III. The simulation results of the different delay budgets and the mixed traffic scenarios are depicted in Section IV, while some conclusions are drawn in Section V.

II. Scheduling Algorithms Formulation

Two sets of algorithms were considered: QoS-differentiated and non-QoS-differentiated algorithms. The latter cannot differentiate between services or QoS demands for each specific user.

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 total priority over scheduling of new data.

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

A. Non-QoS-differentiated algorithms

1) Round Robin (RR): The users are served in a cyclic order ignoring the channel conditions [8]. 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 favourable radio conditions in providing a high instantaneous throughput relative to their average throughput [8]. The priority value for user i is calculated based on Equation 1:

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

where R_i is the estimated bit rate for the next transmission attempt and T_i is the average supported throughput. This average is calculated using a simple exponential smoothing filter [8].

3) Queue Based Max CIR (QBMC): This algorithm incorporates to the classic Max Carrier-to-Interference Ratio (CIR) algorithm, information regarding the queue size of each user's transmission buffer [8], [9]. 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:

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

where CIR_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 Delay Scheduler (LDS) and Asymptotic Delay Scheduler (ADS): The concept of this scheduler was proposed by [2] 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:

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

where BF_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:

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

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

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 inspired by the Modified Largest Weighted Delay First (M-LWDF) algorithm described in [10], [8].

The values assumed for D_{th} are based on each service class requirement. 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.

In this work, the delay threshold assumed by the scheduler is the same as the adopted by the discard mechanism at the MAC-hs (see Section III). 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.

2) Weighted Proportional Fair (WPF): The WPF algorithm works almost the same way as the classical PF scheduler. The only difference is a fixed multiplicative weight, W_{s_i} , which is a QoS differentiation factor to be considered for the service class s_i of user i, as can be seen in Equation 6. This is a simple way to establish a priority hierarchy between different service classes.

$$p_i = W_{s_i} . \frac{R_i}{T_i} \, . \tag{6}$$

The weight values assumed for VoIP and web browsing services are 2.0 and 1.0, respectively.

III. Simulation Modeling

The present research made use of a discrete time system-level dynamic simulator that models the forward link of a Release 99/Release 5 WCDMA system. This section comprises the computational models used in this software tool.

A. HSDPA

The most important aspects related to HSDPA were modeled in the simulations, such as: Adaptive Modulation and Coding (AMC) based on a maximum throughput rule and the amount of data available in the MAC-hs buffer; HARQ soft combining (Chase Combining); HARQ with parallel Stop-And-Wait (SAW) processes; MAC-hs Service Data Unit (SDU) discard mechanism; code multiplexing where the base station transmission power available for HSDPA will be shared equally among all the channelization codes (physical channels) of the multiplexed users.

When a MAC-hs transport block is transmitted on the HS-DSCH in a 2 ms Transmission Time Interval (TTI), the corresponding Block Error Probability (BLEP) is read from the Average Value Interface (AVI) look-up tables that depend on the instantaneous channel quality, the modulation and coding scheme, and the channel profile.

B. Traffic models

The Web Browsing and voice traffic models considered in this study are modeled according to [11]. These services use Radio Link Control (RLC) Acknowledged Mode (AM) and Unacknowledged Mode (UM), respectively.

3rd. Generation Partnership Project (3GPP) has chosen the Adaptive Multirate (AMR) codec to be used in the VoIP service. During ON periods, with AMR mode 12.2 kbps, the VoIP application generates 32-byte voice payload at 20 ms intervals. It is assumed that only one AMR packet is encapsulated in the Real-Time Transport Protocol (RTP) packet every 20 ms. According to 3GPP specifications, the MAC-d SDU payload must have 39 bytes which is appropriate for VoIP service with header compression. Thus, it is assumed that the total average protocol overhead including RTP, User Datagram Protocol (UDP), Internet Protocol (IP), Packet Data Convergence Protocol (PDCP) and RLC is composed of 7 bytes.

C. VoIP delay budget

According to [4], 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 end-to-end delay should account for all the nodes in the communication path. The present research, is interested in 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 [3], [4] considered delay budgets inside UTRAN in the range from 80 ms to 150 ms. This range should be sufficient for scenarios where the VoIP call is between two mobiles or between a land-line and a mobile user. This work considered a fixed delay budget of 150 ms in the simulations.

To compensate for variations in delay, the receiving terminal employs a play-out buffer. This buffer might discard packets that arrive too late (packet deadline). The deadline is the upper bound of the tolerable delay budget.

D. Radio propagation

Detailed radio-propagation models including distance attenuation [11], spatial correlated shadow fading and Additive White Gaussian Noise (AWGN) multi-path propagation characteristics are incorporated.

E. Radio resource management

The Associated Dedicated Physical Channel (A-DPCH) is power controlled and can be in soft handover mode. Admission control based on the power not used by HSDPA (dedicated and common channels) is used for A-DPCH channels. These three Radio Resource Management (RRM) algorithms are modeled with great level of details.

F. Performance metrics

A VoIP user is assumed as satisfied if it is not blocked and 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). A VoIP packet loss can occur in three different ways: wireless channel errors, discard at the MAC-hs layer and discard at the play-out buffer. A Web data user is regarded as satisfied if it is not blocked and its session throughput is higher or equal to 64 kbps.

The system capacity (offered load) will be represented by the estimated total number of users of all service classes (VoIP and Web) per minute in each cell (sector). This estimative considers the Poisson arrival rate and the mean session duration of each service class.

The system capacity regions are defined as the set of expected number of users (offered load) for which acceptable system-level quality is sustained for all service classes (90% of user satisfaction for both VoIP and Web). The capacity region is constructed varying the traffic mix among the considered service classes, including single service evaluations.

IV. Simulation Results

The simulation parameters are shown in Table I. The reader should focus on the relative comparisons presented in the following graphics, since the absolute results depend highly on the specific parameters defining the scenario and are generally only illustrative.

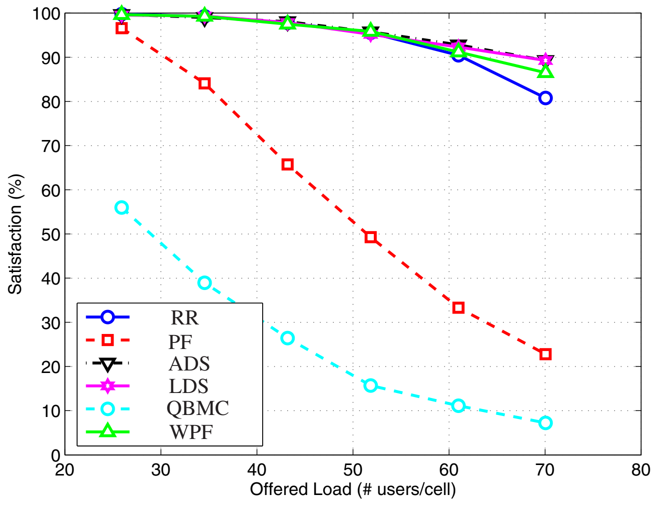
TABLE I
SIMULATION PARAMETERS

Parameter	Value	Unit
Number of cells (torus grid)	27	-
Cell radius	500	m
Maximum BS power	20	W
Power reserved for common channels	3	W
User speed	3	km/h
Number of codes reserved for HSDPA	5	-
Number of H-ARQ parallel processes	6	-
Max. H-ARQ retransmissions	5	-
Average power per HS-SCCH	0.5	W
Max. code multiplexed users	4	-
VoIP satisfaction threshold	90	%
VoIP FER threshold	1	%
Web satisfaction threshold	90	%
Web throughput threshold	64	kbps

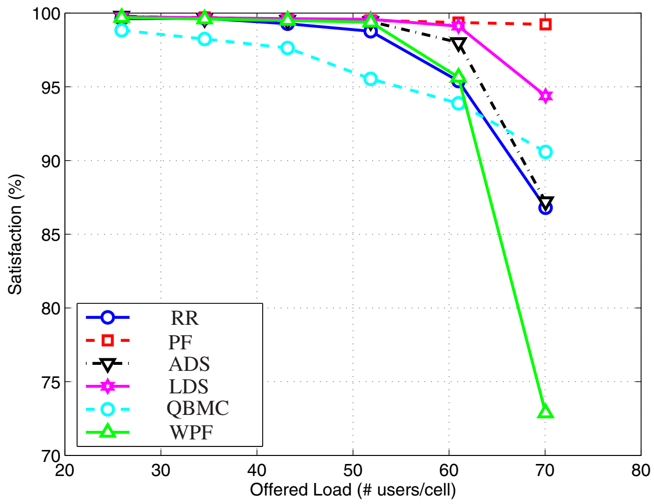
Sections IV-A, IV-B and IV-C present the performance results of the studied PS algorithms for different VoIP delay budgets inside UTRAN (110, 130 and 150 ms, respectively) in a mixed traffic scenario characterized by the same proportion of VoIP and Web users (50% for each). It is important to mention that the discard time threshold at the MAC-hs is assumed to be always 10 ms less than the delay budget, i.e. 100, 120 and 140 ms, respectively. 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 expires. Furthermore, Section IV-D extends the evaluation presented in the previous sections by showing the system capacity regions of all PS algorithms, all traffic mixes (VoIP only, 75% VoIP / 25% Web, 50% VoIP / 50% Web, 25% VoIP / 75% Web, and Web only), and also all VoIP delay budgets considered in this study.

A. Delay budget 110 ms

Figures 1(a) and 1(b) present the network performance for VoIP and Web users, respectively, when the delay budget is set to 110 ms and the MAC-hs discard timer is set to 100 ms. The performance of the VoIP service in Figure 1(a) shows a remarkable degradation of the QBMC algorithm. This is mainly due to the higher buffer sizes experienced by web



(a) VoIP



(b) Web

Fig. 1. User satisfaction ratio with VoIP delay budget of 110 ms and MAC-hs discard timer of 100 ms.

browsing users, resulting in a higher priority to them according to the algorithm formulation in Section II. Both ADS and LDS presents similar behavior. The RR and WPF present a slight poorer performance than these last mentioned algorithms, but much better than the classical PF.

Regarding the Web service, Figure 1(b) presents the PF as the best algorithm followed by the LDS. The ADS and RR algorithms showed almost the same performance for the highest load simulated. The Web satisfaction ratio of the QBMC was not good for low loads, but it outperformed the ADS, RR and WPF algorithms in the highest load. The higher weight given to VoIP considered by the WPF resulted in a bad performance for Web users.

B. Delay budget 130 ms

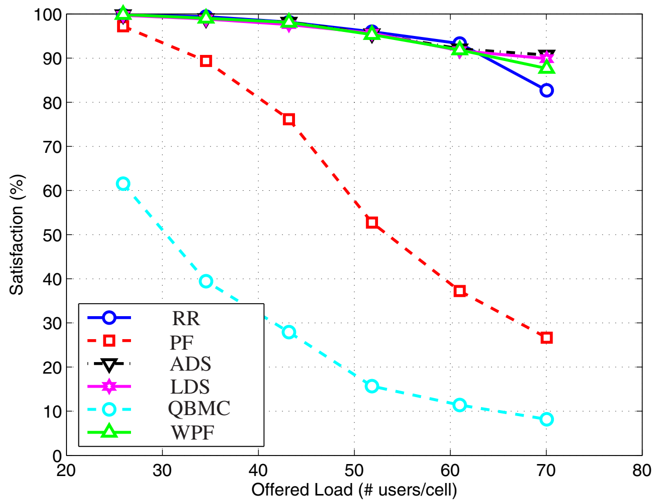
When the delay budget is increased to 130 ms, the performance of VoIP users stays almost constant, as can be seen in Figure 2(a), compared to Figure 1(a). The ADS, LDS and WPF algorithms provide very similar performances, while the QBMC is very uncapable of assuring a good QoS. The RR has a slightly worse performance compared to the best algorithms at the highest simulated load. The PF is much better than the QBMC but still very inferior to the other algorithms. Both of these last algorithms provide a better performance compared to the lower delay budget.

Regarding the Web users, the performance of the PS algorithms depicted in Figure 2(b) remained almost unchanged. The only noticeable change was the improvement of the ADS performance, which became closer to the QBMC performance at the highest load.

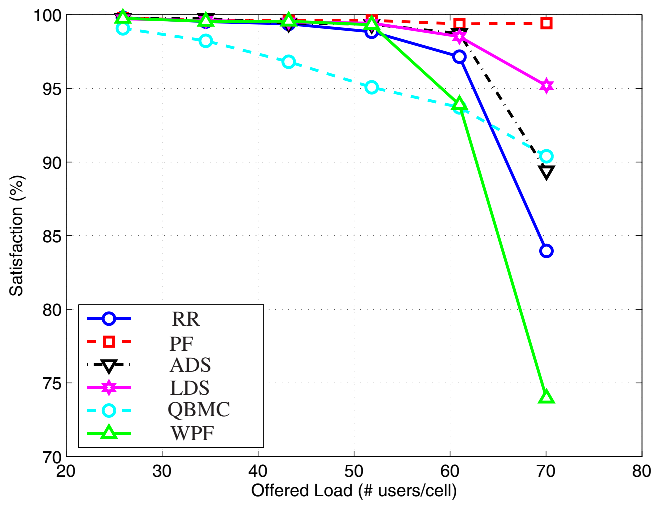
C. Delay budget 150 ms

Now, the system is evaluated in a scenario characterized by the delay budget of 150 ms and MAC-hs discard timer of 140 ms.

The performance of VoIP users is shown in Figure 3(a). An overall satisfaction increase can be noticed, leading to capacity gains. This can be explained by the fact that a



(a) VoIP



(b) Web

Fig. 2. User satisfaction ratio with VoIP delay budget of 130 ms and MAC-hs discard timer of 120 ms.

delay requirement less strict was considered, which causes a decrease in the occurrence of packet losses due to excessive delay. The difference in the performance of ADS, LDS, RR and WPF algorithms became negligible.

The satisfaction ratio regarding Web users is presented in Figure 3(b). The relative behavior of almost all the PS algorithms was the same observed in the lower delay budgets. The only exception was the ADS algorithm, which improved its performance as the delay budget became less strict.

D. System capacity regions

This section presents some capacity figures regarding all the scheduling algorithms and assuming the three VoIP delay budgets considered in this study. The capacity is based on the QoS limit regarding all the mixes, aiming at a detailed definition of how each algorithm is better regarding the mixed services configuration.

Figure 4 presents the capacity regions for all studied PS algorithms when the VoIP delay budget is 110 ms and the MAC-hs discard timer is 100 ms. It can be noticed that the ADS and LDS provide the highest overall capacity, with the former performing better than the latter in the VoIP single service scenario. For a single service scenario comprised of Web users only, the performance of the ADS and LDS is equal to the PF, as expected.

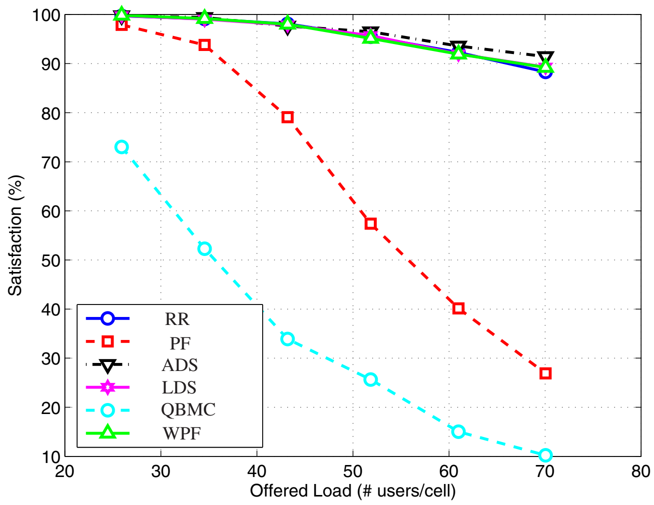
The RR has a good performance when the VoIP traffic is present, but a poor capacity when there is only Web users. The PF has the opposite behavior: very good performance for Web service, but a very bad performance when VoIP is present. The WPF presented a remarkably better performance than the PF and higher capacity than the RR for all mixes.

The QBMC has only a good performance when a single service scenario comprised of only VoIP users is considered. For all the other scenarios, it has the worst performance. This can be explained by the higher buffer sizes experienced by web browsing users, which leads to a higher priority for this service.

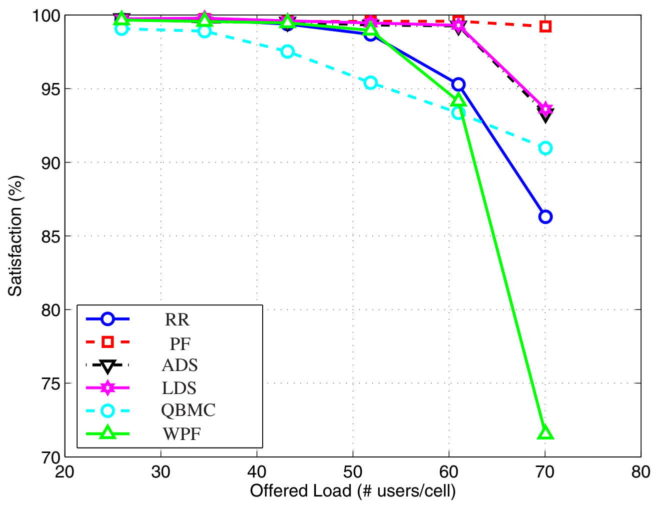
When the VoIP delay budget is increased to 130 ms, there is naturally an improvement in the capacity when the VoIP service is offered in Figure 5. The behavior of the capacity curves is quite similar. The only difference is that the improvement obtained by the ADS is higher.

Figure 6 presents the capacity regions for all studied PS algorithms when the VoIP delay budget is 150 ms. It can be noticed that the ADS provides the highest overall capacity. The advantage is higher when the proportion of VoIP users is increased. For a single service scenario comprised of Web users only, the performance of the ADS is equal to the PF, as expected. It is then able to maximize the capacity for all mixes. The performance of the LDS follows the same behavior as the ADS, but not so efficiently. Observing Figures 4, 5 and 6, one can notice that the higher the delay budget, the more evident is the gain of ADS over LDS.

Another conclusion is that, as the delay budget is increased, the WPF algorithm capacity becomes worse than the RR algorithm when the proportion of VoIP users is higher. This can be explained by the fact that, considering a higher delay



(a) VoIP



(b) Web

Fig. 3. User satisfaction ratio with VoIP delay budget of 150 ms and MAC-hs discard timer of 140 ms.

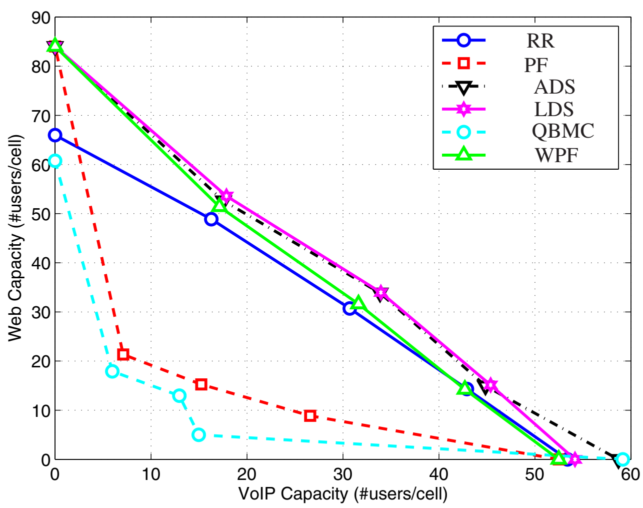


Fig. 4. Capacity regions for all traffic mixes with VoIP delay budget of 110 ms and MAC-hs discard timer of 100 ms.

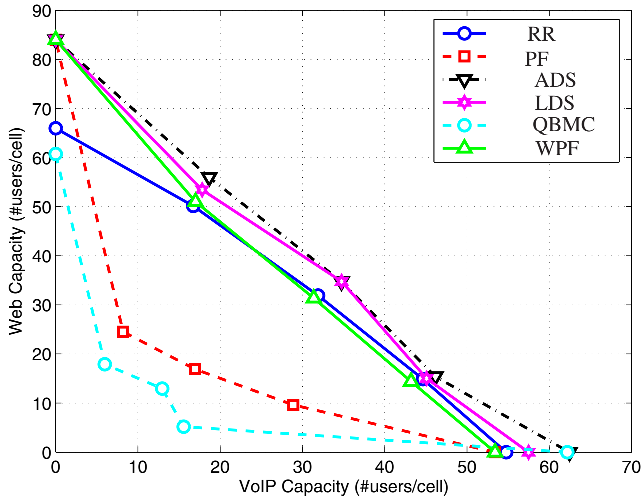


Fig. 5. Capacity regions for all traffic mixes with VoIP delay budget of 130 ms and MAC-hs discard timer of 120 ms.

budget, the capacity limitation in the system is set by the Web service for both schedulers. Decreasing the delay budget leads the system capacity to be limited by the VoIP service. In this last scenario, the WPF algorithm provides more prioritization to VoIP users compared to the RR algorithm.

V. Conclusions

The main contribution of this work was a complete evaluation of the capability of some scheduling algorithms in providing QoS to the VoIP service in a mixed traffic scenario where the resources are shared with Web users. The QoS of the Web service was also considered for the calculation of the capacity limits.

The QBMC provides a good performance when only VoIP users are offered to the system, but has a considerably poor performance when the Web traffic is inserted in the network.

Although the PF algorithm provides very bad results due to limitations on the VoIP service, the algorithms that obtained the better results are some sort of modified PF. When only the VoIP service is offered to the system, the ADS and LDS provide huge capacity gains compared to classical algorithms like the PF and RR. When the proportion of Web traffic is increased in the system, they are still capable of providing good QoS to both Web and VoIP services, leading to the best overall capacity.

The WPF does not provide as good performance as the ADS and LDS, mainly if there is a higher proportion of VoIP users, due to capacity limitation regarding the Web service. However, it works much better than the pure PF, becoming a reasonable and simple solution for services differentiation.

It can be concluded that, in general, the PS algorithms able to perform QoS differentiation presented better performance in terms of capacity and QoS compared to those non-QoS-differentiated algorithms. Among the former category, those that take into consideration delay requirements of the VoIP service (ADS and LDS) presented the best overall capacity for all traffic mixes and delay budgets.

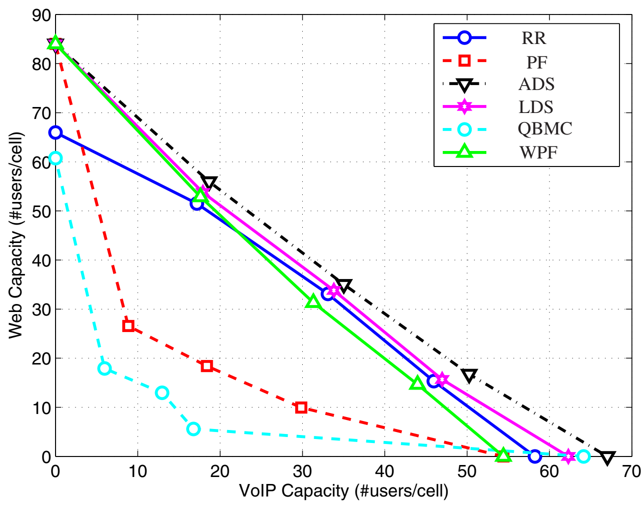


Fig. 6. Capacity regions for all traffic mixes with VoIP delay budget of 150 ms and MAC-hs discard timer of 140 ms.

Acknowledgments

This work was supported by a grant from Ericsson of Brazil - research branch under ERBB/UFC.17 technical cooperation contract. Francisco R. P. Cavalcanti was partly funded by CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico, grant no. 304477/2002-8. The authors would like to thank the valuable comments of Sven-Olof Jonsson, Stefan Wänstedt and Mårten Ericson from Ericsson Research in Luleå, Sweden.

References

[1] A. R. Braga, E. B. Rodrigues, and F. R. P. Cavalcanti, “Novel Scheduling Algorithms Aiming for QoS Guarantees for VoIP over HSDPA,” *VI International Telecommunications Symposium - ITS*, September 2006.

[2] P. Hosein, “Capacity of Packetized Voice Services over Time-Shared Wireless Packet Data Channels,” *IEEE INFOCOM*, March 2005.

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

[4] B. Wang, K. I. Pedersen, T. E. Kolding, and P. E. Mogensen, “Performance of VoIP on HSDPA,” *IEEE 61st Vehicular Technology Conference - VTC Spring*, May 2005.

[5] H. Zheng, G. Rittenhouse, and M. Recchione, “The Performance of Voice over IP over 3G Downlink Shared Packet Channels under Different Delay Budgets,” *IEEE Global Telecommunications Conference - GLOBECOM*, pp. 2501–2505, October 2003.

[6] 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.

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

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

[9] 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 PIMRC*, vol. 3, September 2004, pp. 2202–2206.

[10] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, and P. Whiting, “Providing Quality of Service over a Shared Wireless Link,” *IEEE Communications Magazine*, pp. 150–154, February 2001.

[11] 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.