PACKET SCHEDULING FOR VOIP OVER HSDPA IN MIXED TRAFFIC SCENARIOS

André R. Braga, Emanuel B. Rodrigues and Francisco R. P. Cavalcanti
Wireless Telecommunication Research Group - GTEL
Federal University of Ceará - UFC
Fortaleza, Brazil
{andrerb, emanuel, rodrigo}@gtel.ufc.br

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. Two sets of algorithms were considered: QoS-differentiated and non-QoS-differentiated algorithms. It can be concluded that, in general, the PS 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.

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) (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 [1]. 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 is particularly challenging. This is due to the fact that wireless channels have lower reliability, and time varying signal strength, which may cause severe QoS violations.

Some works have dealt with PS algorithms for VoIP service on High Speed Downlink Packet Access (HSDPA) in a single service scenario [1, 2, 3]. The only work that has assessed the VoIP service in a mixed traffic scenario was [4], but its focus was on 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. The novelty of the present work comes from the fact that it provides a complete evaluation of the VoIP service for several PS algorithms in terms of QoS and capacity, considering mixed traffic scenarios. 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.

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 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.

A. *Non-QoS-differentiated algorithms*

1) *Round Robin (RR)*

The users are served in a cyclic order ignoring the channel conditions [5]. 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 [5]. 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 [5]

1-4244-0330-8/06/$20.00©2006 IEEE

3) Queue Based Max CIR (QBMC)

This algorithm incorporates to the classic Max CIR algorithm, the information regarding the queue size of each user's transmission buffer [5, 6]. With this information, the scheduler is able to prioritize queues that are not being served due to its channel conditions. In fact it uses an indirect information of the delay. The priority calculation of user i is presented in Equation (2):

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

where C_i is the measured CIR and S_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 [1] 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. The priority calculation for each user i is presented in Equation (3):

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

where B_i is the BF for user i, which can be calculated by using Equations (4) or (5):

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

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

considering $D_{th_{s_i}}$ as the delay threshold and D_i is the head-of-line packet delay for user i. When $D_{th_{s_i}}$ is equal or larger than $D_{th_{s_i}}$, a constant value of 100 is assumed for the BF. The first is a variation of the BF proposed by [7] aiming at the possibiliy of scheduling in a mixed service scenario, while the other is an alternative way of calculation the BF considering a linear function. The LDS is inspired in algorithms proposed in [8, 5].

The values assumed for $D_{th_{s_i}}$ are based on each service class requirement. The BF calculation presented above has an advantage of, when a service does not have delay requirements, $D_{th_{s_i}}$ can be set to a large 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. 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):

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

This is a simple way to establish a priority hierarchy between different service classes. Since the speech service is considered to be the most important, 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 the UMTS. This section comprises the computational models used in this software tool.

A. High Speed Downlink Packet Access (HSDPA)

The most important aspects related to HSDPA were modeled in the simulations, such as: Adaptive Modulation and Coding (AMC) based on link conditions and the amount of data available in the MAC-hs buffer (no feedback error assumed); Hybrid Automatic Repeat Request (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 available base station transmission power for HSDPA is equally shared 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 [9]. These services use Radio Link Control (RLC) Acknowledged Mode (AM) and Unacknowledged Mode (UM), respectively.

The Adaptive Multirate (AMR) 12.2 kbps vocoder generates a voice frame every 20 ms during activity periods (an active factor of 0.5 is considered). According to the 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 all the protocol layers is composed of 7 bytes.

C. VoIP delay budget

According to [3], 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 present research is interested at the delay budget inside UMTS Terrestrial Radio Access Network (UTRAN). This delay budget should be enough for all the the Node B functionalities (scheduling, HARQ procedures, etc.) and the user reception of VoIP packets. Studies conducted in [2, 3] 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 [9], 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 is based on the power not used by HSDPA (dedicated and common channels), also called the non-HS power.

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 average packet 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) 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 1. The results are separated regarding the service proportion. The first part of the evaluations concerns a service mix where there the VoIP dominates in number of users (75% VoIP / 25% Web). The other results are composed of both services in equal proportions (50% VoIP / 50% Web). The last scenario is comprised of a domination of the Web service (25% VoIP / 75% Web). The last presented result summarizes the capacity limits of all traffic mixes.

Table 1: 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
MAC-hs discard timer (VoIP)	140	ms
MAC-hs discard timer (Web)	infinite	ms
VoIP delay budget	150	ms
VoIP satisfaction threshold	90	%
VoIP FER threshold	1	%
Web satisfaction threshold	90	%
Web throughput threshold	64	kbps
PF filter coefficient	0.00125	-

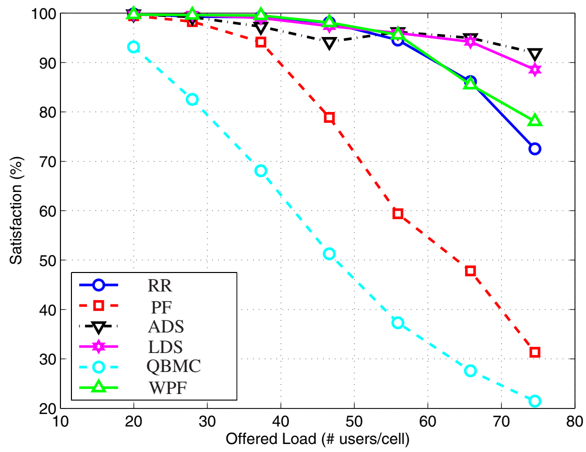
A. 75% *VoIP* + 25% *Web*

Figures 1(a) and 1(b) present the network performance for VoIP and Web users, respectively. The performance of the VoIP service shows a remarkable degradation of the QBMC algorithm. This is mainly due to the higher buffer sizes experienced by web 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, the PF turns out to be the best algorithm followed by the QBMC. The higher weight given to VoIP considered by the WPF results in a bad performance for Web users. The ADS provides better results than both the LDS and RR. It also presents a slight poorer performance compared to both the PF and QBMC algorithms under high loads. In summary, the algorithms that provide a better performance to one service have the worse performance regarding the other, due to service priorization (ADS, LDS, WPF and QBMC) or due to resource distribution (RR and PF).

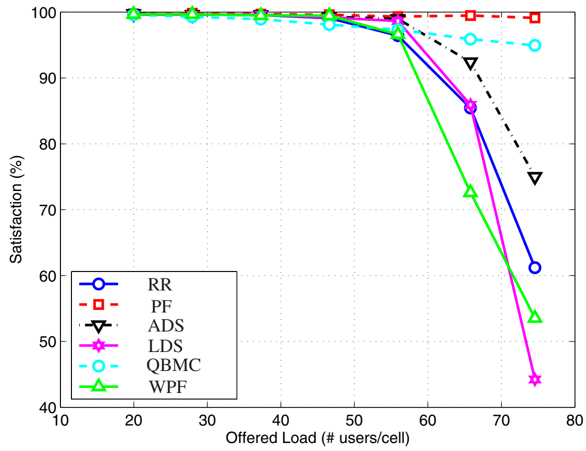
B. 50% *VoIP* + 50% *Web*

The performance of VoIP users is presented in Figure 2(a). It is clear that the QBMC has the worst quality provided, followed by the PF. The relative behavior of both algorithms remains the same as the previous mix (see Figure 1(a)) but the absolute performance is worse. On the other hand, the RR and WPF improved their performance. Both the ADS and LDS obtained the same good performance.

The satisfaction ratio regarding Web users is presented in Figure 2(b). The superiority of the PF is not so evident as

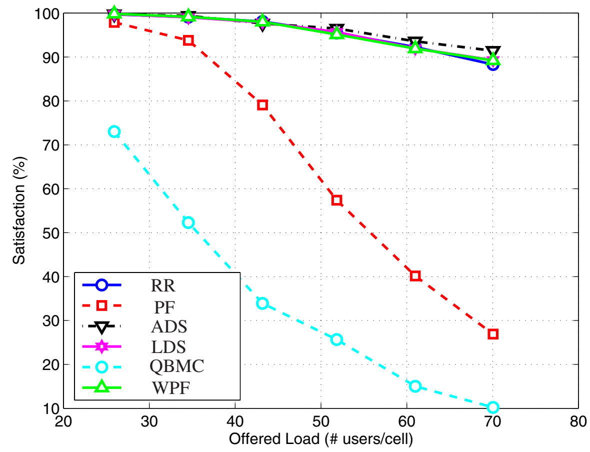


(a) VoIP

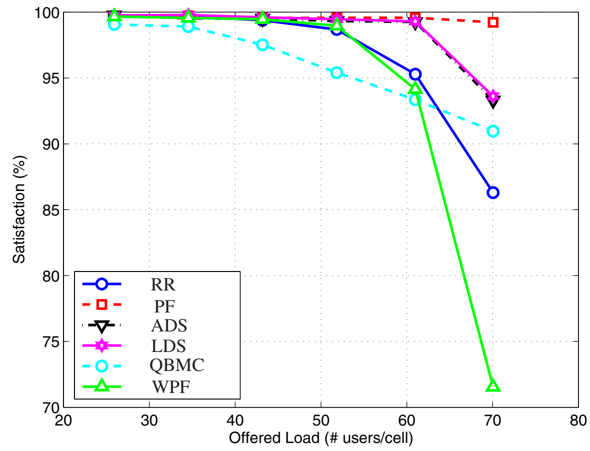


(b) Web

Figure 1: User satisfaction ratio (75% VoIP, 25% Web).



(a) VoIP



(b) Web

Figure 2: User satisfaction ratio (50% VoIP, 50% Web).

the obtained with a smaller proportion of the Web service in the offered load. Both the ADS and LDS provide very close performance. The RR and WPF provide quite similar performance, with a slight advantage (about 1% more satisfied users) for the RR under overload situations.

It is clear that, except for the case of QBMC and PF, there is a significant improvement of all other algorithms in the performance of the Web service compared to the previous mix (see Figure 1(b)). This can be explained by the fact that these algorithms provide a service priorization to VoIP. Since there is a less amount of VoIP users, there is more resource left to the Web service.

C. 25% *VoIP* + 75% *Web*

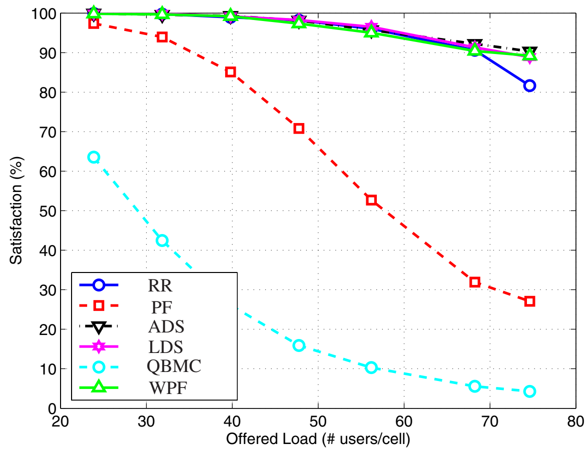
The performance of VoIP users is shown in Figure 3(a). There is a similar behavior compared to the scenario with equal number of VoIP and Web users (see Figure 2(a)). The ADS, LDS and WPF algorithms provide very similar performances, while the QBMC is very uncapable of assuring a good QoS. The RR has 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 worse performance compared to the previous mix.

One interesting fact concerning this scenario regards the WPF scheduler. It is capable of providing a reasonable performance for both services. It can be explained that, since there is a less amount of high priority users in the system, the scheduler is able to perform a resource reservation to it and also provide a good QoS to the other service (Web). The performance of the WPF for web users outperforms now the RR, as can be seen in Figure 3(b).

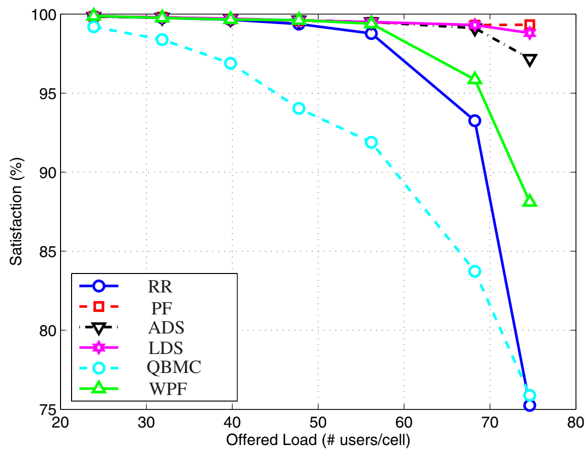
D. *System capacity regions*

Figure 4 presents the capacity regions for all studied PS algorithms. It can be noticed that the ADS provides the highest overall capacity. The advantage is higher when the proportion of VoIP users is increased. 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. 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 has a remarkably better performance than the PF.



(a) VoIP



(b) Web

Figure 3: User satisfaction ratio (25% VoIP, 75% Web).

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.

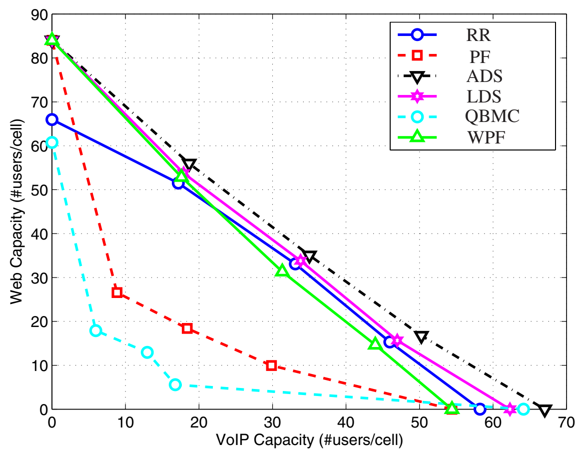


Figure 4: Capacity regions for all traffic mixes.

V. Conclusions

The QBMC provides a good performance when only VoIP users are offered to the system (only worse than ADS), but has a considerable poor performance when the Web traffic is inserted in the network. Regarding the RR and WPF, the former showed better results in scenarios with equal proportion of services and dominated by VoIP users, while the latter performed well in scenarios with majority of Web users. Moreover, WPF works much better than the pure PF in all traffic mixes, 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.

Acknowledgement

This work was supported by a grant from Ericsson of Brazil - research branch under ERBB/UFC.17 technical cooperation contract.

References

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

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

[3] 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.

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

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.