The 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'07)

LOAD CONTROL FOR VOIP OVER HSDPA IN MIXED TRAFFIC SCENARIOS

Emanuel B. Rodrigues, Francisco Rafael M. Lima and Francisco Rodrigo P. Cavalcanti
Wireless Telecommunications Research Group - GTEL / Federal University of Ceará - UFC
Fortaleza, Brazil

ABSTRACT

By means of dynamic system level simulations, this work evaluates the performance of two proposed Load Control algorithms that are capable of guaranteeing the QoS of the VoIP service in a mixed traffic scenario in the HSDPA network. It was shown that the LC algorithms are efficient at maximizing the VoIP satisfaction, and therefore the overall system capacity, by means of a soft and controlled degradation of the QoS of the WWW service.

I. INTRODUCTION

Recently, there has been a growing interest in providing Real Time (RT) services, e.g. Voice over IP (VoIP), in wireless packet data networks, such as High Speed Downlink Packet Access (HSDPA) in the all-IP concept. This would bring improved efficiency in terms of resource sharing, spectrum usage and network architecture. On the other hand, providing VoIP service is a challenging task due to its intrinsic restrictions. In this context, VoIP-oriented Load Control (LC) algorithms are used to control the system load in order to guarantee Quality of Service (QoS) and increase the VoIP capacity by means of the degradation of lower priority services.

Conventional LC algorithms used to be dimensioned for the worst case conditions, given by the network planning, and were designed to monitor whether the load reached a fixed target load. Due to the limitations of conventional LC schemes, some research has been done in adaptive LC [3, 5]. In the adaptive LC, the target load is not fixed, but is rather dynamically adapted according to the system status. This approach brings advantages because a fixed target load cannot track a complex time-variant environment such as wireless cellular systems.

In this work, we propose two adaptive LC algorithms for HSDPA in a mixed service scenario. They are composed by two Radio Resource Management (RRM) functionalities: an admission control algorithm, called Session Admission Control (SAC) [1], and a scheduling algorithm named Weighted Proportional Fair (WPF) [2]. The two proposals differ basically in how the priority of VoIP sessions are dynamically adapted and which VoIP performance metric is used to detect VoIP overload.

The remainder of this paper is organized as follows. Section II presents the mathematical formulation of all the evaluated LC 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. LOAD CONTROL ALGORITHM FORMULATION

A brief introduction of the SAC and WPF techniques is presented in sections A and B, while the formulation of the proposed LC algorithms is shown in section C.

A. *Session Admission Control*

The SAC is a tool which aims to provide QoS guarantees to sessions with high priority by denying admission to flows with lower priority, once a system overload situation in the high priority sessions is detected.

In our work, the SAC scheme is employed to guarantee VoIP QoS in a mix with World Wide Web (WWW) sessions. The SAC scheme considers delay as the resource to be shared among users in the system. Therefore, the VoIP packet delays are regularly measured and filtered at the Node B and then reported to the Radio Network Controller (RNC). At RNC, there are two admission thresholds depending on the type of session, which are compared to the filtered measurement reported from the Node B. The difference between VoIP and WWW admission thresholds is called SAC priority margin. Therefore, in a session admission event, the admission controller verifies if the current VoIP delay measurement reported from the Node B, added to the estimated resource usage of the incomming session, is higher or lower than the admission threshold for that type of session. If higher, the access of the incomming session is blocked, otherwise the incomming session is submitted to a Link Admission Control (LAC) which would verify power and code availability. For more details about SAC see [1].

B. *Weighted Proportional Fair*

The WPF scheduler is based on the Proportional Fair (PF) scheduling algorithm and provides a fixed priority for each service class. The priority value for this scheduler is given by $p_i = W_{s_i} \cdot \left(\frac{R_i}{T_i}\right)$; where W_{s_i} is a multiplicative weight for the service class s_i of user i used for QoS differentiation, R_i is the estimated bit rate for the next transmission attempt and T_i is the average supported throughput. For more details about the WPF scheduler see [2].

The WPF scheduler can improve the VoIP capacity through the assignment of a higher weight factor for VoIP compared to WWW. In our work, the VoIP weight factor is fixed to 1 and the WWW weight factor is given by the difference between the VoIP weight factor and a WPF priority margin. The WPF priority margin is a non-negative number in the interval $[0, 1]$, which measures the level of priority degradation of WWW compared to VoIP, i.e., the higher the WPF priority margin, the higher the VoIP priority over WWW sessions.

1-4244-1144-0/07/$25.00©2007 IEEE

C. *Load Control*

The strategy of assigning non-zero SAC and WPF priority margins gives precedence to VoIP sessions in the High Speed Downlink Shared Channel (HS-DSCH) access and can improve the VoIP QoS through the degradation of the WWW session QoS. Therefore, in order to guarantee the VoIP QoS, the proposed LC algorithms will adjust in a controlled and adaptive manner the WWW QoS by changing the priority margins between these two services in the SAC and WPF schemes. These priority margins must be understood as the priority degradation margins of the WWW service compared to the VoIP service.

Let's define α and β as the priority degradation margins of the WWW service compared to VoIP in the SAC and WPF strategies, respectively. In the LC algorithms formulation, α and β are assumed to be in dB scale and constrained to the range of [-10, 0] dB. For example, consider that the SAC admission threshold and the WPF priority weight for the VoIP service are $D_{VoIP}^{th} = 150$ ms and $W_{VoIP}^{prio} = 1$, respectively. If $\alpha = \beta = -0.5$ dB, then the SAC admission threshold for the WWW service is given by $D_{WWW}^{th} = D_{VoIP}^{th} \cdot 10^{\frac{\alpha}{10}} \simeq 133.69$ ms and the WPF priority weight for the WWW service is given by $W_{WWW}^{prio} = W_{VoIP}^{prio} \cdot 10^{\frac{\beta}{10}} \simeq 0.89$.

Two LC algorithms are proposed in this work: **Jump Based Load Control (JLC)** and **Error Feedback Based Load Control (EFLC)**. Both of them are inspired by the Wideband CDMA (WCDMA) Outer-loop Power Control (OLPC) framework. Table 1 shows the similarities between the OLPC and the adaptive downlink LC algorithm.

The main objective of the LC algorithms is to make sure that the Frame Erasure Rate (FER) of the VoIP users connected to a given sector is around a planned value. VoIP frames of an user can be lost in two different ways: errors caused by the wireless channel (air interface) and packet discard at the play-out buffer due to unnacceptable delay (higher than the VoIP delay budget). The VoIP QoS is closely related to the profile of service prioritization. Thus, the outer loop of the LC algorithms monitors the VoIP QoS in the sector regularly. If it observes that the VoIP QoS is not being fulfilled, it will change the target values of the outer loop resource, which are the priority margins between the VoIP and WWW services. In the proposed LC algorithms, two target values are updated dynamically: the SAC priority margin (α) and the WPF priority margin (β). If the VoIP QoS in the sector is worse than a desired target, α and β will be updated in order to degrade WWW QoS and direct more resources to the VoIP service. On the other hand, if the VoIP QoS is excessively good, more priority can be given to the WWW service, so that the sector resources are used more efficiently. The SAC and WPF algorithms act as important players in the inner loop of the LC algorithm. By means of their actions (accept/refuse WWW connections or schedule WWW users more/less frequently), they have indirect control of the inner loop resource, which is the delay experienced by the VoIP users. This inner loop control tries to fulfill the service prioritization target that was chosen by the outer loop.

Table 1: Comparison between OLPC and the proposed LC.

	OLPC	**LC**
Desired quality	BLER	VoIP FER
Dynamic Output	SINR target	SAC prio. margin (α); WPF prio. margin (β)
Resource in outer loop	SINR	Service prioritization
Resource in inner loop	Tx power	VoIP delay

The time basis for the adaptation of the α and β parameters must be aligned with the UMTS Terrestrial Radio Access Network (UTRAN) architecture constraints. Since the SAC algorithm is executed at the RNC, the α parameter must be calculated at the Node B and reported to the RNC using the Iub interface and the Node B Application Part (NBAP) signaling. It is reasonable to think that a good time period for the calculation and reporting of the α parameter would be 100 ms, which is the same reporting time of other measurements in the NBAP signaling. Regarding the β parameter, there is no need to report it to the RNC, since the packet scheduling algorithm is executed at the Node B. The β parameter can be calculated every 2 ms, which is the HSDPA Transmission Time Interval (TTI) and also the time basis for the fast packet scheduling.

In the next sections, the two LC algorithms proposed in this study are described.

1) *Jump Based Load Control Algorithm*

The proposal of the JLC algorithm was inspired from the algorithm presented in [3]. The parameters α and β in the JLC are updated according to the well-known OLPC jump algorithm proposed in [4].

In the proposed JLC algorithm, the trigger of the jump algorithm will be a VoIP QoS outage event. An outage event occurs when the VoIP packet delay averaged over all users connected to a given sector is higher than the delay budget of the VoIP service. More details about the VoIP delay budget can be found in section III. The algorithm is described below.

1. Check for a VoIP QoS outage event during the last time window, which depends on the LC actuation frequency.

2. If outage is true then

$$\alpha(t) = min\{max\{\alpha_{min}, \alpha(t-1) - \Delta\}, \alpha_{max}\} \quad (1)$$

else

$$\alpha(t) = min\{max\{\alpha_{min}, \alpha(t-1) + \Delta/K\}, \alpha_{max}\} \quad (2)$$

The β parameter is adapted in the same way as the α parameter given in (1) and (2). Notice that in the β equations, β_{min} and β_{max} take place. In (1) and (2), Δ is the step size of the JLC algorithm in dB. $K \geq 1$ is an integer that is related to the jump in the target values of α and β when there was not an outage event in the last time window. If we configure $K = \frac{1}{FER_{VoIP}^{target}} - 1$, we guarantee that the VoIP FER will be always lower or equal to FER_{VoIP}^{target} [4]. α_{min}, α_{max}, β_{min} and β_{max} are the minimum and maximum values in dB of the α and β parameters, respectively.

2) *Error Feedback Based Load Control Algorithm*

The proposal of the EFLC algorithm was based on the work developed in [5]. An interesting idea from [5] was the dynamic adaptation of a power threshold in the LAC algorithm based on the packet loss rate of the IP Radio Access Network (RAN). This adaptation used a feedback control strategy based on a constrained integral control law. In the EFLC algorithm proposed in the present work, the VoIP FER is monitored by each sector. Since all downlink VoIP traffic will be scheduled in the Node B, the sector can calculate the FER averaged over all the VoIP users connected to it. The priority margins α and β are calculated periodically in the sector by comparing the monitored VoIP FER with a target VoIP FER value. With the VoIP FER measured and filtered in the last control interval (time window of algorithm actuation) represented as $FER_{VoIP}^{filt}(t)$ and the target VoIP FER value represented as FER_{VoIP}^{target}, the new parameter α is calculated as follows:

$$\alpha(t) = min\left\{max\left\{\alpha_{min}, \alpha(t-1) - \sigma_\alpha \cdot e(t)\right\}, \alpha_{max}\right\} \tag{3}$$

$$e(t) = FER_{VoIP}^{filt}(t) - FER_{VoIP}^{target} \tag{4}$$

The β parameter is adapted in the same way as the α parameter given in (3). Notice that in the β equation, β_{min}, β_{max} and σ_β take place. The parameters σ_α and σ_β control the adaptation speed of the parameters α and β, respectively. α_{min}, α_{max}, β_{min} and β_{max} are the minimum and maximum values in dB of the α and β parameters, respectively. $e(t)$ given by (4) is the control error, which is the difference between the monitored variable (FER_{VoIP}^{filt}) and its target (setpoint) value.

For a constrained integral controller, high σ_α and σ_β values lead to a faster response, but care is required since high values can cause oscillations or instabilities. After studying the controller convergence, we assumed $\sigma_\alpha = \sigma_\beta = \frac{\Delta}{1-FER_{VoIP}^{target}}$, where Δ is the step size of the JLC algorithm in dB.

Paper [5] claims that it is possible to obtain fast reactivity of the integral controller without introducing high variance by adjusting the controlled parameter in a nonlinear (e.g., exponential) fashion. Based on that, it was decided to use a filtered value of the VoIP FER ($FER_{VoIP}^{filt}(t)$) and use it in (4) and (3) (the same applies to β). The time series $FER_{VoIP}(t)$ was filtered using a Simple Exponential Smoothing (SES) filter, which is a first order Infinite Impulse Response (IIR) filter suitable for time series with slowly varying trends. The SES method also has a forecasting property since it learns from the past errors: the estimate for period $t+1$ is increased if the current value for period t is greater than what was estimated to be and decreased otherwise. The relative influence of recent and older data, i.e. the actual filter tuning, is regulated by the filter constant.

The main differences between the JLC and EFLC algorithms are the step size for the adaptation of the α and β parameters, and the way the algorithms decide whether the VoIP QoS requirement was met or not. On one hand, the JLC uses fixed step sizes (step-up and step-down) depending if a discrete VoIP QoS outage based on delay has occured or not. On the other hand, the EFLC uses a dynamic step size whose calculation is directly derived from the VoIP QoS metric that is being targeted (FER). Thus, it is expected that EFLC is able to perform a more fine-tuned control of the VoIP FER towards the desired value.

III. Simulation Modeling

The present research made use of a discrete time system-level dynamic simulator that models the forward link of the WCDMA R99/R5 system. This section comprises the most important computational models used in this software tool.

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) Chase Combining; HARQ Stop-And-Wait (SAW) processes; 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 TTI, the corresponding Block Error Probability (BLEP) is read from the Average Value Interface (AVI) look-up tables that depend on the channel quality, the modulation and coding scheme, and the channel profile.

The WWW and VoIP traffic models considered in this study are modeled according to [6]. The VoIP traffic model creates packets that mimic the Adaptive Multirate (AMR) 12.2 kbps codec. DTX packets are not considered.

According to [2], VoIP delay budgets inside UTRAN in the range from 80 ms to 150 ms are acceptable. This work considered a fixed VoIP delay budget of 150 ms in the simulations. To compensate for variations in VoIP delay, the receiving terminal employs a play-out buffer. This buffer might discard VoIP packets that arrive too late (packet deadline). The deadline is the upper bound of the tolerable VoIP delay budget.

Detailed radio-propagation models are incorporated in the simulator, such as distance attenuation [6], spatial correlated shadow fading and single-path Rayleigh small-scale fading.

The Associated Dedicated Physical Channel (A-DPCH) is power controlled and can be in soft handover mode. LAC for A-DPCH is based on the power not used by HSDPA (dedicated and common channels), also called the non-HS power.

A VoIP user is assumed as satisfied if it is not blocked and has a FER lower or equal to 1% in the downlink, reflecting a good perceived speech quality provided by the AMR codec with 2% end-to-end FER. A WWW data user is regarded as satisfied if it is not blocked and its average 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 WWW) in each cell/sector. This estimate 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 WWW). 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 2. The results presented in this section concern the satisfaction ratio of both VoIP and WWW services, where the performance of the proposed LC algorithms is compared with a reference scenario. In the reference scenario, no load control algorithm was configured, the SAC was configured with a fixed priority margin of $0\ ms$, and the WPF priority margin was 0, i.e., the PF scheduler was configured (see sections A and B for the definition of SAC and WPF priority margins). The objective in this scenario is to remove service prioritization between VoIP and WWW given in the admission and scheduling procedures.

The service mixes considered for VoIP and WWW users are: 25% / 75%, 50% / 50% and 75% / 25%. The last set of evaluations summarizes the capacity limits of all traffic mixes, including single service scenarios.

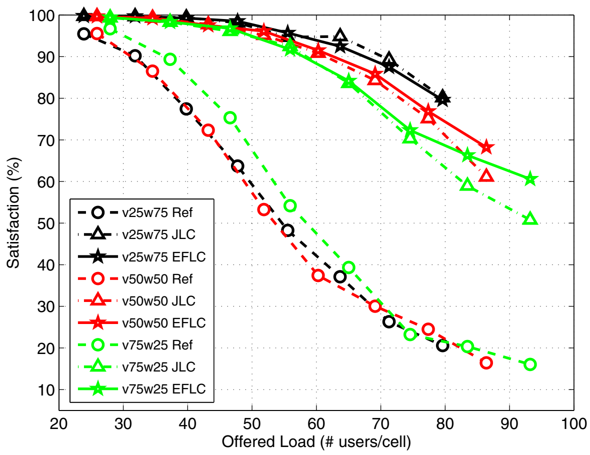
Figs. 1(a) and 1(b) clearly show that the proposed LC algorithms are efficient at maximizing the VoIP satisfaction compared to the reference scenario for all the traffic mixes considered in this work. Among the LC algorithms, it can be seen that the higher the proportion of VoIP users in the traffic mix, the lower the satisfaction of both service classes. More VoIP users means that the LC algorithms have a harder work to do in this more challenging network scenario: guarantee the QoS of a larger number of highly demanding VoIP users, and softly pre-empt network resources of fewer WWW users.

It can be seen in Figs. 1(a) and 1(b) that as the proportion of VoIP users in the traffic mix increases, the higher is the gain of the EFLC algorithm over the JLC algorithm regarding the VoIP satisfaction. The advantages of the EFLC compared to the JLC were the reasons for that difference in performance, such as variable LC step size, and consequently, fine-tuning control of the LC parameters α and β; more quickness when leading with congestion situations due to the forecasting property of the exponential filtering; and the better synchronized action of the SAC and WPF schemes. At low and moderate offered loads, the JLC and EFLC strategies mostly use the adaptive WPF scheme to control the delay experienced by the VoIP packets. However, at high offered loads, the SAC scheme in the EFLC was stricter than the JLC. The former started earlier to prevent a huge number of WWW users from entering the system in order to guarantee the QoS of the ongoing VoIP connections.

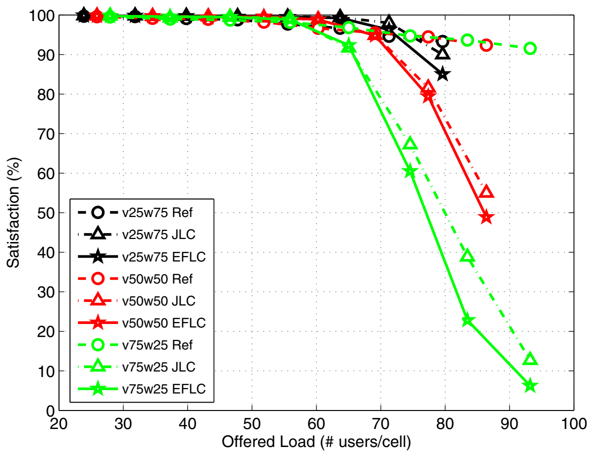
Fig. 2 presents the capacity regions regarding all the LC algorithms and the reference scenario. The LC algorithms

Table 2: 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
VoIP delay budget	150	ms
VoIP SAC delay threshold	150	ms
VoIP WPF priority weight	1	-
VoIP FER target	1	%
SAC actuation period	100	ms
WPF actuation period	2	ms
Maximum value of α and β	0	dB
Minimum value of α and β	-10	dB
JLC step size	0.5	dB
JLC delay target	150	ms
EFLC exponential filter coefficient	0.1	-
VoIP and WWW satisfaction threshold	90	%
VoIP FER threshold (satisfaction)	1	%
WWW throughput threshold (satisfaction)	64	kbps



(a) VoIP



(b) WWW

Figure 1: User satisfaction ratio for different traffic mixes.

perform a service based QoS control. Looking at the reference scenario depicted in Figs. 1(a) and 1(b), and considering a satisfaction threshold of 90% for both service classes, one can see that the overall system capacity was strongly limited by the VoIP service, while the WWW QoS was excessively good. The LC algorithms performed a smooth and controlled degradation of the WWW QoS in order to free network resources and maintain the VoIP FER around the planned value, providing a considerable increase in the VoIP satisfaction. Although the VoIP continues to be the limiting service, the overall system capacity, which is defined as the minimum capacity between the service classes, is maximized. This statement can be proved in Fig. 2, where one can see the larger capacity region provided by the LC strategies when compared to the reference scenario. Moreover, the JLC and EFLC presented very similar performance around the 90% satisfaction threshold (within statistical confidence interval). However, as shown in Fig. 1(a), the EFLC is more successful than the JLC at guaranteeing the VoIP QoS for wider ranges of offered traffic loads and traffic mixes.

Fig. 3 shows the relative capacity gains (global and for each service class) achieved by the EFLC algorithm over the reference scenario for all traffic mixes. It can be observed that the EFLC algorithm degraded the QoS of the WWW service in order to protect the QoS of the VoIP service, and consequently maximized its capacity. The more VoIP users that exist in the system, the lower the VoIP capacity gain, and the higher the WWW capacity loss. It is important to notice that since the VoIP was the capacity limiting service for all the traffic mixes, the overall system capacity gain was equal to the VoIP capacity gain. The EFLC algorithm presented a global capacity gain from 60% to 110% depending on the traffic mix considered.

V. Conclusions

The EFLC algorithm presented equal or better VoIP satisfaction than the JLC algorithm for all the ranges of traffic loads and traffic mixes considered in the simulations. This was due to the advantages of the EFLC over the JLC, such as:

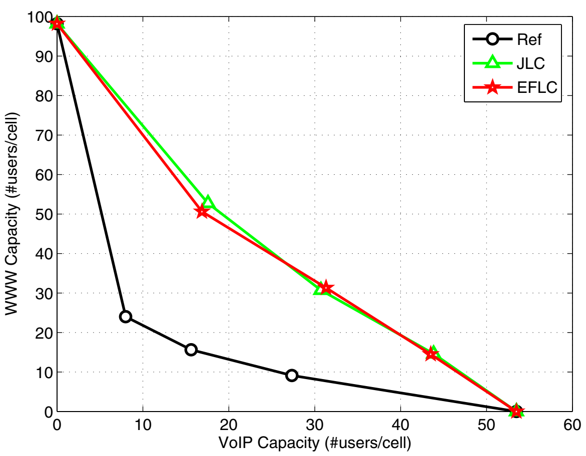


Figure 2: System capacity regions for different traffic mixes.

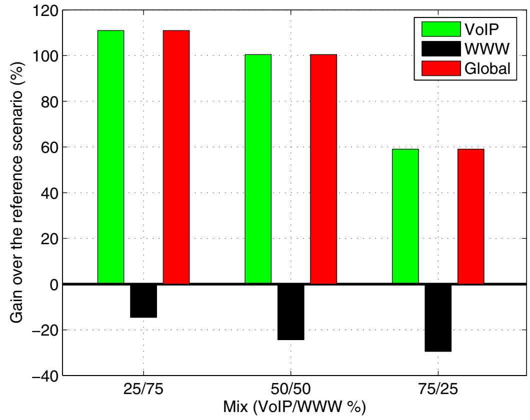


Figure 3: Global and individual capacity gains of the EFLC over the reference scenario.

variable load control step size, and consequently, fine-tuning control of the LC parameters α and β, which are related to the SAC and WPF priority margins, respectively; more quickness when leading with congestion situations due to the forecasting property of the exponential filtering; and the better synchronized action of the SAC and WPF schemes.

Furthermore, the EFLC strategy presented significant overall capacity gains compared to the reference scenario. The lowest percentual capacity gain was 60% for the traffic mix 75% VoIP + 25% WWW, while the highest percentual capacity gain was 110% for the traffic mix 25% VoIP + 75% WWW.

Acknowledgement

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

References

[1] A. R. Braga, S. Wänstedt and M. Ericson, "Admission Control for VoIP over HSDPA in a Mixed Traffic Scenario," *VI IEEE International Telecommunications Symposium - ITS*, pp.397–402, September 2006.

[2] 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.

[3] J. Gu and X. Che, "Adaptive Uplink Load Control in CDMA Systems," *62nd IEEE Vehicular Technology Conference - VTC Fall*, vol. 2, pp.1193–1196, 2005.

[4] A. Sampath, P. S. Kumar and J. M. Holtzman, "On Setting Reverse Link Target SIR in a CDMA System," *47th IEEE Vehicular Technology Conference - VTC Spring*, vol. 2, pp.929–933, May 1997.

[5] S. K. Kasera, R. Ramjee, S. R. Thuel and X. Wang, "Congestion Control Policies for IP-Based CDMA Radio Access Networks," *IEEE Transactions on Mobile Computing*, vol. 4, n. 4, pp.349–362, 2005.

[6] UMTS/ETSI, "Selection procedures for the choice of radio transmission technologies of the UMTS," *UMTS TR 101.112 v.3.2.0*, April 1998.