(Re)Active Load Control Based on Radio Link Quality for the UMTS/WCDMA Forward Link

Emanuel B. Rodrigues, Francisco R. P. Cavalcanti, Carlos H. M. de Lima and Vicente A. de Sousa Jr.
Teleinformatic Engineering Department
Federal University of Ceará (UFC)
Fortaleza-CE, Brazil
Email: {emanuel, rodrigo, carlos, vicente}@gtel.ufc.br

Abstract— **This contribution specifies congestion resolution (action) and recovery (reaction) procedures within a Load Control (LC) functionality in the WCDMA downlink air interface. It is verified by means of dynamic system level simulations that, in soft overload situations, it is advantageous to give lower priority to those radio links that have worse channel quality, so that the system can return to its stable state in a gradual and controlled manner.**

I. INTRODUCTION

In the UMTS/WCDMA forward link, it is of utmost importance to maintain the air interface load below thresholds defined in the network planning phase. The reason behind this is that an excessive load prevents the network from guaranteeing the contracted requirements. The planned coverage area is not provided, the capacity is lower than required and the Quality of Service (QoS) is degraded. Moreover, an air interface load above the allowed value can lead the network into an unstable condition.

Several factors can cause overload situations in the forward link of the cell. The external interference is a statistical factor that depends on the mobility pattern and user service requirements, and therefore, cannot be controlled. The sector transmission power to different users depends on their locations on the cell coverage area. Unusual localization patterns (most of the users are far from the BS antenna, for example), can demand an amount of power unavailable in the sector at that moment. The existence of a network utilization pattern in specific periods of the day is a major factor that influences on the presence of congestion.

The soft overload is a very important aspect of WCDMA systems. The soft overload provides a compromise between system stability, number of voice calls being served by the system and the received voice quality (conversation intelligibility). An increase of the blocking rate and a decrease of the voice intelligibility quality for a short time is assumed to be an acceptable behavior for a system in overload situations. In other words, the system degrades the quality of service in a gradual and controlled fashion in congestion situations, so that the ongoing calls and the system stability are guaranteed.

This work dealt with the development of procedures of the Load Control (LC) algorithm (action and reaction) using a coherent combination of other RRM algorithms functionalities, such as Call Admission Control (CAC) and Rate Adaptation (RA) via Adaptive Multi-Rate (AMR) mode selection.

In contrast to the congestion problem in fixed networks, where a large number of studies have been published, and despite radio network congestion being a widely recognized and identified problem, not many specific solutions and algorithms in a wireless environment are available in the open literature [1-4] and few of them [5-9] are well aligned to 3GPP specifications, as is the case for the algorithms proposed and studied in this contribution.

In the references [6, 7], the Congestion Control technique is studied in the reverse link of the WCDMA system. The forward link is contemplated in [9, 8]. However, the former considers a load factor based congestion control, while the latter do not use a dynamic system level simulator to evaluate the LC algorithm performance. A downlink transmission power based congestion control is evaluated in [5], but the study was focused on the interactive service class only.

Therefore, there is a clear contribution of the research described in this work. Not only is the work strongly related to practical applications of the UMTS system but it evaluates the Load Control algorithm based on the downlink transmission power by means of dynamic simulations as well. Different priority selection criteria and rate adaptation for the conversational service class is considered.

The remainder of this paper is organized as follows. Section II presents the simulation models considered in this research. The congestion resolution and recovery procedures are described in section III, while the performance results taken from the dynamic simulations are depicted in section IV. Finally, we draw the main conclusions of the work in section V.

II. SYSTEM LEVEL DYNAMIC SIMULATION MODELING

The simulation environment is an hexagonal multi-cell deployment with uniformly distributed mobile terminals. The simulated system consisted of 16 cell sites, each serving three hexagonal sectors. A wrap around technique is used to avoid border effects in interference calculations.

Voice call requisitions are generated according to a Poisson process. We model an ON-OFF traffic pattern, with activity and silent periods being generated by an exponential distribution.

0-7803-8887-9/05/$20.00 (c)2005 IEEE

The macrocell test environment that was considered is the Vehicular Test Environment with low-speed users at 3 km/h. The propagation effects considered are mean path loss, slow variation in the signal due to shadowing, and rapid variation due to multipath effects and scattering. The impact of multipath propagation on channel fading, downlink orthogonality loss and RAKE receiver performance are included in the simulations. Our power and handover control strategies are strongly based on the 3GPP standards.

III. CONGESTION RESOLUTION AND RECOVERY PROCEDURES

In general, the Load Control algorithm functionality operates in a TTI basis (20 ms) and is divided in two phases: congestion resolution and congestion recovery. The start and end of these phases are determined by a congestion detection procedure [10, 7, 6]. The algorithm formulation is described below.

1) **Congestion Detection**: A decision criterion based on the sector active transmission power must decide whether the cell should initiate or not a congestion resolution/recovery stage. If the sector transmission power stays above a threshold $\mho_{cong}^{resol}$, whose value is relative to the maximum transmission power available for traffic channels and was defined in the network planning phase, during a time window of ΔT_{cong}^{recov}, the Load Control algorithm supposes that the cell is congested and must initiate a phase to resolve the overload. After the actions needed to reduce the cell load, if the sector transmission power stays below a threshold $\mho_{cong}^{recov}$, whose value is relative to the maximum transmission power available for traffic channels and was defined in the network planning phase, during a time window of ΔT_{cong}^{recov}, the algorithm initiates a new stage in order to recover from the congestion situation.
2) **Congestion Resolution**: An algorithm based on the following three steps can be used in order to guarantee the network stability:
    a) **Priority Order**: Ordination of the different users from the lowest priority to the highest, obeying a pre-established criterion, which can be based on random choice or radio link quality.
    b) **Load Reduction**: Several actions in order to reduce the cell load in congestion periods can be performed by means of the interaction between the Load Control algorithm and other RRM algorithms (power control, handover control, call admission control and rate adaptation).
    c) **Load Check**: The actions to reduce the load must be performed until the sector transmission power is lower than a threshold defined in the network planning phase. This threshold is the same used by the congestion detection procedure to decide the beginning of the recovery stage. If the congestion persists, the algorithm have to return to step b) and contemplate the next users in the priority list.
3) **Congestion Recovery**: In this phase, it is possible to restore the standard configurations of the RRM algorithms and the transmission parameters of the users by means of a scheduling process following the inverse order of the priority list adopted in the congestion resolution phase.

The time windows used to detect overload situations are sliding windows with duration ΔT_{cong}^{resol} and ΔT_{cong}^{recov}, relative to the decision of the resolution and recovery actions, respectively. A percentile of 80% is considered in the distribution of transmission power samples of the sectors collected every Transmission Time Interval (TTI) of 20 ms within the observation window. For instance, consider an observation window of 200 ms. Considering a TTI of 20 ms, 10 samples of the mean sector transmission power will be collected. At least 8 (80%) of the samples must satisfy the criterion defined to trigger the congestion resolution and recovery processes.

As mentioned previously, the Load Control algorithm interacts with other RRM algorithms to solve an overload situation in the cell. In order to re-establish the cell stability by means of load reduction in the congestion resolution phase, the Load Control (LC) commands the Call Admission Control (CAC) to reject any admission of new calls or connections coming from other cells (handover). Furthermore, the Rate Adaptation (RA) algorithm is used to reduce the transmission powers of the traffic channels by means of mode selection of the AMR voice codec. Only one user will have its AMR mode decreased in each TTI (20 ms). Forced dropping of connections is not allowed in the simulations.

The Adaptive Multi-Rate (AMR) voice codec has eight operation modes (source rates): 12.2, 10.2, 7.95, 7.40, 6.70, 5.90, 5.15 and 4.75 kbps. Due to the lack of typical E_b/N_0 requirement values in the literature for all the different AMR modes, the research have used only three of them: AMR-12.2 (12.2 kbps), AMR-7.95 (7.95 kbps) and AMR-4.75 (4.75 kbps). The AMR codec is capable of changing its transmission rate dynamically at each voice frame of 20 ms by means of commands of the radio access network. Due to the different E_b/N_0 requirements considered for each AMR mode (7 dB for AMR-12.2, 7.5 dB for AMR-7.95 and 8 dB for AMR-4.75) [11], inferior AMR modes (AMR-7.95 and AMR-4.75) require less transmission power than the AMR-12.2 mode. In this way, the total transmission power of the sector can be decreased so that the congestion problem is solved. Afterwards, in the congestion recovery phase, the AMR modes used by the mobile stations before the congestion detection will be restored and the call admission will be liberated. Again, only one user will have its AMR mode increased in each TTI (20 ms).

In this work, it is assumed that the voice users are satisfied if they experience a Frame Erasure Probability (FEP) lower than 2%, no matter which AMR modes are assigned to them. It is known that different AMR modes present distinct voice inteligibility patterns. However, this aspect is not included in the performance evaluation. It is considered that the lowest AMR mode (4.75 kbps) is still able to provide a reasonable voice quality, and therefore, the most important metric regarding the user satisfaction is the FEP.

Fig. 1 depicts the LC algorithm procedure evaluated in the simulations. $\overline{P_{on,r}^{ED,Tx}}$ is the mean transmission power of the active traffic channels of a generic sector r; $P_{max,r}^{ED,Tx}$ is the maximum cell transmission power reserved for traffic channels (16.5 W); $\mho_{cong}^{resol}$ and $\mho_{cong}^{recov}$ are detection thresholds of congestion resolution and recovery, respectively; ΔT_{cong}^{resol} and ΔT_{cong}^{recov} are the observation windows for detection of congestion resolution and recovery, respectively.

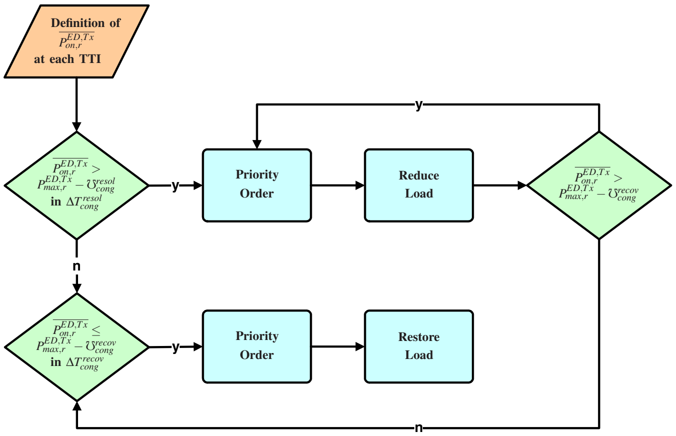

Fig. 1 Load control based on the mean transmission power of a sector.

IV. SIMULATION RESULTS

As indicated in section III, the congestion resolution phase is composed of three procedures: priority order definition, load reduction and load check. The congestion recovery phase is responsible for restoring the transmission parameters (AMR modes) that the users were using before the load reduction in the congestion resolution phase take place. The reduction and restoring of the transmission parameters follow a priority order. The users will be classified based on a criterion that will define who will be the first to have his transmission data rate decreased (resolution) or increased (recovery). In the resolution stage, the priority ordination is performed from the lowest to the highest priority. In the recovery phase, the inverse order is adopted. Three priority definition criteria are proposed in this research:

1) **Best** E_c/N_0 - The radio link quality is characterized by the received chip energy relative to the total power spectral density (E_c/N_0) of the Common Pilot Channel (CPICH) on the forward link. The users who have the best propagation channel quality have the lowest priority.
2) **Worst** E_c/N_0 - In this case, the lowest priority is allocated to those users who present the worst CPICH E_c/N_0, and consequently the worst propagation conditions.
3) ***First In First Out*** (**FIFO**) - In this criterion, the first users that were admitted in the system have lower priority. They will be the first ones to have their AMR mode decreased in the congestion resolution phase and last ones to have their AMR mode restored during the congestion recovery phase. In comparison to the other two criteria, the FIFO strategy can be regarded as a random choice, since we do not know the radio link quality of the oldest users in the system.

The CPICH E_c/N_0 measurement is reported from the mobile station to the base station every TTI (20 ms). The same reporting framework used for handover purposes can be used in the congestion control algorithm.

The simulation parameters related to the LC algorithm are defined as follows: the measurement period of the sector transmission power is 667 μs, the filtering window of the sector transmission power measurements is 20 ms, the observation window for detection of both congestion resolution (ΔT_{cong}^{resol}) and recovery (ΔT_{cong}^{recov}) is 100 ms, and the detection threshold of congestion resolution ($\mho_{cong}^{resol}$) and recovery ($\mho_{cong}^{recov}$) are -0.5 dB and -1 dB, respectively.

The Reference Scenario considered in the following performance results do not use any Load Control functionality. However, a basic Admission Control algorithm is employed in all cases (scenarios with/without LC algorithm).

The main performance indicator of the LC algorithm is the system congestion rate, which is the percentage of time that the cells are considered congested based on the criterion used to determine the beginning of the resolution phase. This metric is presented in fig. 2. It can be seen that the load reduction procedure of the resolution phase worked properly, since the three cases that used the load control technique presented lower congestion levels compared to the Reference Scenario, which do not adopt any LC functionality. As expected, the FIFO criterion turned out to be a compromise between the two other criteria (best and worst CPICH E_c/N_0). The effect of the load control on the power consumption of the sectors can be visualized in fig. 3, which depicts the 95^{th} percentile of the normalized active transmission power utilization of the cells (maximum value = 20 W).

The blocking probability is strongly influenced by the portion of time that the system remains in congestion situations. This happens because we assume that a congested cell will deny any new admission requisition or handover connections coming from other cells. However, the LC algorithm acts in favor of the users that had been already admitted to the system (ongoing calls), because the QoS requirements of these users are threatened by the overload situation. Even when a congested cell denies access to a handover connection, the Load Control algorithm do not drop this call. Instead of that, the user remains connected to the same sector. It is assumed that in this special case (overload situation), the user would accept to possibly experiment a degraded channel quality instead of being dropped. A weighted function of the blocking and dropping probabilities is depicted in fig. 4 (Grade of Service - GoS). The load control functionality is able to provide better quality of service in terms of *GoS* compared to the Reference Scenario, which do not use any technique to control the congestion.

The QoS perceived by the users and the transmission power allocation of the sectors can be visualized indirectly by means of the rejection rate of the power control TPC commands. When the sector is very congested and there is not any transmission power available, the base station is capable of denying power control commands $TPC = 1$ (power increase) from the mobile station. This metric can indicate that the load control is able to control the system coherently in order to avoid instability situations, such as the overconsumption of the sector transmission power. This performance indicator is illustrated in fig. 5. It can be seen that the Reference Scenario presents higher probability of denying power increment commands from the power control algorithm, since the LC algorithm is not used in this scenario. This fact indicates that the Reference Scenario remains congested for longer periods of time.

To conclude the user quality of service analysis, fig. 6 depicts the 95^{th} percentile of the Frame Erasure Probability (FEP). It can be observed that the LC algorithm is capable of guaranteeing the users' QoS requirements even for high offered loads. Furthermore, as indicated by the other performance indicators, the selection criterion that gives lower priority to those users that present the worst CPICH E_c/N_0 obtained the best results. This fact suggests that during the load reduction procedure of the congestion resolution phase, it is advantageous to decrease the data rate (lower AMR mode) of those users that experiment poor radio link quality. This priority order is inverted in the congestion recovery stage, where the original AMR modes must be restored to each user that had been selected in the resolution phase.

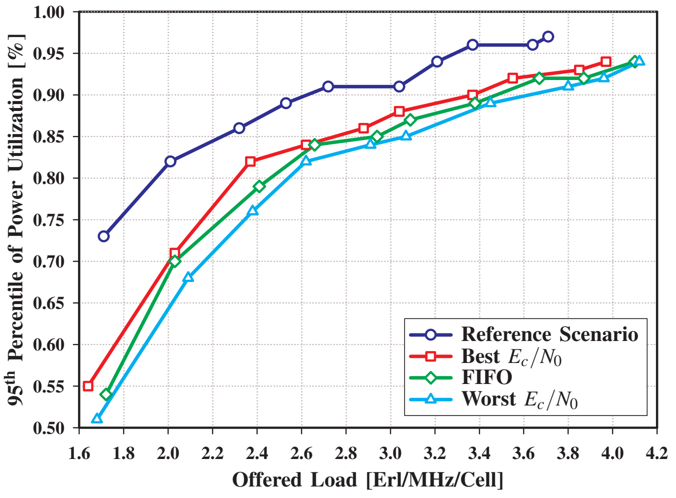

Fig. 3 95^{th} percentile of the normalized active transmission power utilization of the sector (maximum value = 20 W) with/without load control considering different priority selection criteria.

V. CONCLUSIONS

The scenarios that utilized the LC algorithm presented lower congestion rate and lower rejection rate of power control commands than the Reference Scenario, which do not use any technique to deal with overload situations.

The presence of the load control functionality caused an increase on the percentage of blocked users; however, the continuous monitoring and regulation of the cell load benefited the existing users, decreasing the dropping probability and

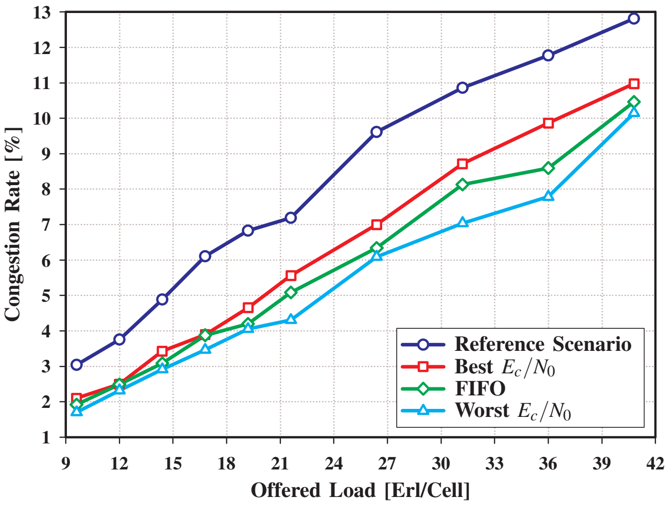

Fig. 2 System congestion rate with/without load control considering different priority selection criteria.

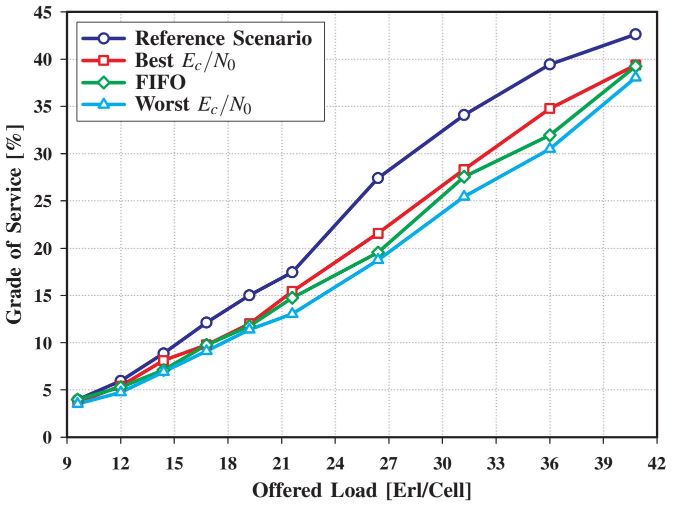

Fig. 4 Grade of Service (GoS) with/without load control considering different priority selection criteria.

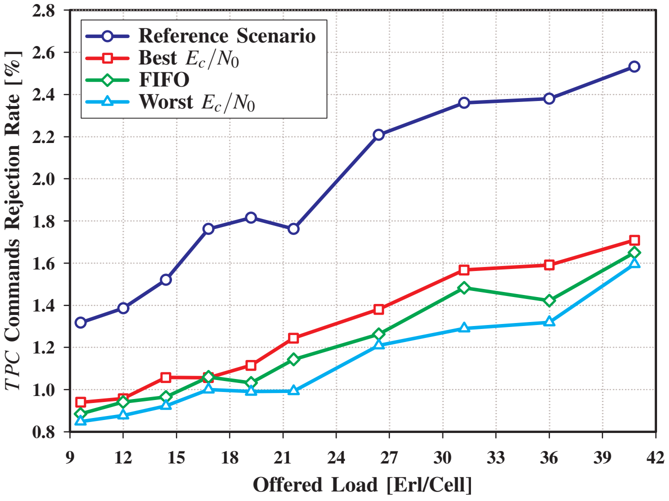

Fig. 5 Rejection rate of the power control TPC commands with/without load control considering different priority selection criteria.

improving the GoS metric in comparison with the Reference Scenario.

It was verified that the LC algorithm is capable of improving the QoS metrics (GoS e FEP) even when a range of high offered loads is considered.

Finally, the selection criterion that gives lower priority to those users that present the worst CPICH E_c/N_0 obtained the best results. This fact suggests that the actions of reduction/restoration of the transmission parameters (AMR mode) in the resolution/recovery phases, should follow a priority order based on this criterion to adapt the data rate of the voice calls via AMR mode selection. However, this criterion yields a high amount of uplink signalling load. The network operator must evaluate whether this higher complexity is worthy. The FIFO criterion seems to be a trade-off, with low complexity and performance close to that presented by the worst CPICH E_c/N_0 criterion.

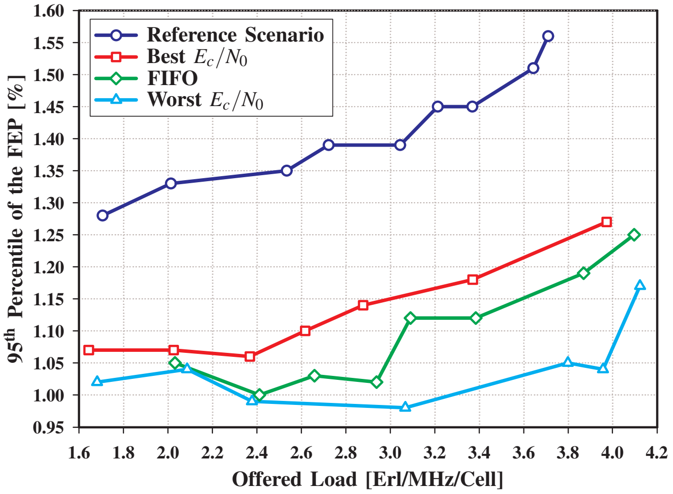

Fig. 6 95^{th} percentile of the Frame Erasure Probability (FEP) with/without load control considering different priority selection criteria.

This work can be extended by the evaluation of a mixed services scenario (conversational and interactive service classes). The Rate Adaptation functionality can be upgraded to control the data packet transmission of non-real-time services by means of Transport Block Set (TBS) reconfiguration. The TBS size can be adjusted every TTI so that less interference is generated in the air interface. In this way, the total transmission power of the sector can be decreased so that the congestion problem is solved.

ACKNOWLEDGMENT

Ericsson of Brazil - Research Branch under ERBB/UFC.07 and ERBB/UFC.08 Technical Cooperation Contracts; CAPES and FUNCAP for scholarship support.

REFERENCES

[1] T. K. Liu and J. A. Silvester, "Joint Admission/Congestion Control for Wireless CDMA Systems Supporting Integrated Services," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 6, pp. 845–857, August 1998.

[2] N. Passas and L. Merakos, "A Graceful Degradation Method for Congestion Control in Wireless Personal Communication Networks," *46th IEEE Vehicular Technology Conference - VTC Spring*, vol. 1, pp. 126–130, May 1996.

[3] F. Baccelli, B. Blaszczyszyn, and F. Tournois, "Downlink Admission/Congestion Control and Maximal Load in CDMA Networks," *22nd Annual Joint Conference of the IEEE Computer and Communications Societies - INFOCOM*, vol. 1, pp. 723–733, March 2003.

[4] A. Abrardo, G. Giambene, and D. Sennati, "Capacity Evaluation of a Mixed-Traffic WCDMA System in the Presence of Load Control," *IEEE Transactions on Vehicular Technology*, vol. 52, no. 3, pp. 490–501, May 2003.

[5] W. Rave, T. Köhler, J. Voigt, and G. Fettweis, "Evaluation of Load Control Strategies in an UTRA/FDD Network," *53rd IEEE Vehicular Technology Conference - VTC Spring*, vol. 4, pp. 2710–2714, May 2001.

[6] J. Pérez-Romero, O. Sallent, R. Agustí, and J. Sánchez, "Managing Radio Network Congestion in UTRA-FDD," *Electronics Letters*, vol. 38, no. 22, pp. 1384–1386, October 2002.

[7] O. Sallent, J. Pérez-Romero, R. Agustí, and F. Casadevall, "Provisioning Multimedia Wireless Networks for Better QoS: RRM Strategies for 3G W-CDMA," *IEEE Communications Magazine*, vol. 41, no. 2, pp. 100–106, February 2003.

[8] D. Soldani and J. Laiho, "A Virtual Time Simulator for Studying QoS Management Functions in UTRAN," *58th IEEE Vehicular Technology Conference - VTC Fall*, October 2003.

[9] J. Mückenheim and U. Bernhard, "A Framework for Load Control in 3rd Generation CDMA Networks," *IEEE Global Telecommunications Conference - GLOBECOM*, vol. 6, pp. 3738–3742, November 2001.

[10] 3GPP, "Radio Resource Management Strategies," 3rd Generation Partnership Project, Sophia Antipolis, France, Tech. Rep. TR 25.922 v6.0.1 - Release 6, April 2004.

[11] H. Holma and A. Toskala, *WCDMA for UMTS - Radio Access for Third Generation Mobile Communications*, 2nd ed. John Wiley & Sons, Ltd, 2002.