Mixing Conversational and Interactive Traffic in the UMTS Radio Access Network

J. Sánchez, J. Pérez-Romero, O. Sallent, R. Agustí.
Universitat Politècnica de Catalunya (UPC), c/ Jordi Girona 1-3, 08034 Barcelona, Spain
Phone: +34 93 401 71 95 Fax: +34 93 401 72 00
email: [jorperez, sallent, ramon @ tsc.upc.es]

Abstract - **The definition and assessment of suitable Radio Resource Management (RRM) strategies able to provide QoS in the framework of the UTRA segment of UMTS is a key issue for achieving the expectations created on 3G technology. This paper proposes and evaluates specific algorithms for the different RRM functions involved in the uplink direction in a scenario with a mixture of interactive and conversational services. In particular the effect of prioritization of conversational users in the admission control has been analyzed in terms of admission, dropping probabilities and packet delay. Furthermore, the paper also studies the importance of suitable congestion control mechanisms that cope with load fluctuations in order to guarantee the negotiated QoS to already connected users. These fluctuations are mainly due to the randomness in the traffic generation of interactive users, that can seriously degrade performance of both conversational and even interactive users if no congestion control is carried out.**

***Keywords*:** RRM; UTRA; admission control; congestion control; interactive services; conversational services.

1. INTRODUCTION

W-CDMA access networks, such as the considered in UTRA-FDD proposal [1], provide an inherent flexibility to handle the provision of future 3G mobile multimedia services. The optimization of capacity in the air interface is carried out by means of efficient algorithms for Radio Resource Management that take into account the average and peak interference levels present in the system [2][3]. These functionalities cover admission control, congestion control and short term RRM to decide suitable transport formats and power levels. Although these functionalities are very important in the framework of 3G systems because they are the basis to guarantee a certain target Quality of Service (QoS), not much effort has been devoted to them up to date in the open literature, especially when all of them are jointly considered. Within this context, this paper presents new admission and congestion control mechanisms for the uplink. The paper is organized as follows: Section 2 details the uplink RRM approach, which is evaluated through system level simulation in Section 4 following the simulation model defined in Section 3. Finally, Section 5 summarizes the obtained results.

2. RRM ALGORITHMS

UMTS radio interface provides a layered architecture where logical channels are mapped to transport channels in the MAC (Medium Access Control) layer. A transport channel defines the way how traffic from logical channels is processed and sent to the physical layer. The smallest entity of traffic that can be transmitted through a transport channel is a Transport Block (TB). Once in a certain period of time, called Transmission Time Interval (TTI), a given number of TB will be delivered to the physical layer in order to introduce some coding characteristics, interleaving and rate matching to the radio frame. The set of specific attributes are referred as the Transport Format (TF) of the considered transport channel. Note that the different number of TB transmitted in a TTI indicates that different bit rates are associated to different TF. The network assigns the list of allowed TF to be used by the UE in what is referred as Transport Format Set (TFS). The configuration of all these parameters is a task of RRM.

3GPP approach for uplink RRM can be divided in two parts:

1. Centralized component (located at RNC). Admission and congestion control algorithms are carried out.
2. Decentralized component (located at UE-MAC). This algorithm autonomously decides a TF within the allowed TFS for each TTI, and thus operates at a "short" term in order to take full advantage of the time varying conditions. In this paper the SCr algorithm based on service credits detailed in [4] is considered.

2.1 *Admission Control*

The admission control procedure is used to decide whether to accept or reject a new connection depending on the interference (or load) it adds to the existing connections. Therefore, it is responsible for deciding whether a new RAB (Radio Access Bearer) can be set-up and which is its allowed TFS. Admission control principles make use of the load factor η and the estimate of the load increase that the establishment of the bearer request would cause in the radio network [5].

From the implementation point of view, admission control policies can be divided into modeling-based and measurement-based policies [6]. In case the air interface load factor η is estimated in statistical terms and assuming that K users are already admitted in a cell, the (K+1)th request should verify:

$$\eta = (1+f)\sum_{i=1}^{K}\frac{1}{\dfrac{SF_i}{v_i\cdot\left(\dfrac{E_b}{N_o}\right)_i\cdot r}+1} + (1+f)\frac{1}{\dfrac{SF_{K+1}}{v_{K+1}\cdot\left(\dfrac{E_b}{N_o}\right)_{K+1}\cdot r}+1} \le \eta_{\max} \quad (1)$$

where SF_i is the spreading factor, v_i the activity factor of the traffic source which is known for both real and non-real time-users depending on traffic characterisation, (Eb/No)i is the i-th user requirement, r the channel code rate and other-cell interference power is modeled as a fraction f of the own-cell received power. In case for example conversational and interactive services are present, two different possibilities are considered:

1.- *Admission without prioritization*: The admission control algorithm does not take into account the class of service of the user who is asking for a connection. For K1 conversational users and K2 interactive users Eq.(1) is applied in all cases in the same way and simply K=K1+K2.

2.- *Admission with prioritization*: The admission principle is to accept a conversational request on the expense, if necessary, of interactive traffic. That is, for an interactive request, Eq.(1) is checked because it is assumed that some kind of "soft QoS" in terms of delay or bit rate (it seems reasonable for example to fix some desired parameters for a web browsing session) should be provided although in strict sense this service class is of non real time nature and, consequently, it is delay tolerant. The admission control for a conversational request will also check Eq.(1). If Eq.(1) holds, the request is accepted. If Eq.(1) does not hold, the congestion control mechanism may be triggered, depending on the real instantaneous load, in order to reduce interactive load and provide room for this request. The setup delay involved in this operation (some hundreds of milliseconds) can be negligible for human perception. Notice that, since the number of users whose TFS need to be reconfigured can be known directly, all of them can be simultaneously reconfigured. Only if not enough capacity can be released from interactive traffic the conversational request is rejected.

2.2 *Congestion Control*

Conventional congestion occurs when the admitted users cannot be satisfied with the normal agreed services for a given percentage of time because of an overload. In this paper, congestion can also be triggered in order to allow the access of a conversational user to the system. The congestion control mechanisms include the following parts:

1) *Congestion detection*: Some criterion must be introduced in order to decide whether the network is in congestion or not. A possible criterion to detect when the system has entered the congestion situation and trigger the congestion resolution algorithm is when the load factor increases over a certain threshold $\eta \ge \eta_{CD}$ during a certain amount of time, ΔT_{CD}.

2) *Congestion resolution*: When a situation of congestion is assumed in the network, some actions must be taken in order to maintain the network stability. The congestion resolution algorithm executes a set of rules to lead the system out of the congestion status. A lot of possibilities exist to carry out this procedure. In any case, three steps are identified:

a) Prioritization: Ordering the different users from lower to higher priority (i.e., from those that expect the lowest degree of service to those with the highest stringent QoS requirements) in a prioritization table. In this paper, all the interactive users are assumed to have the same requirements. Therefore, the criterion consists in giving lower priority to the users transmitting with a higher instantaneous bit rate. So the users will be ordered from lower to higher spreading factor.

b) Load reduction: Two main actions can be taken:

b1) No new connections are accepted while in congestion

b2) Reducing the TFS (i.e. limiting the maximum transmission rate) for a certain number of users already accepted in the network, beginning from the top of the prioritization table. In this case, we assume that these users are not allowed to transmit any more while in congestion period (i.e. the TFS is limited to TF0). This is carried out through the layer 3 RRC (Radio Resource Control) protocol message "Transport Channel Reconfiguration".

c) Load check: After the actions taken in b), one would check again the conditions that triggered the congestion status. If congestion persists, one would go back to b) for the following group of users in the prioritization table. It is considered that the overload situation has been overcome if, for a certain amount of time ΔT_{CR} the load factor is below a given threshold $\eta \leq \eta_{CR}$.

3) *Congestion recovery*: A congestion recovery algorithm is needed in order to restore to the different mobiles the transmission capabilities they had before the congestion was triggered. It is worth mentioning that such an algorithm is crucial because depending on how the recovery is carried out the system could fall again in congestion. The operation of the algorithm consists in increasing progressively the TF on a user by user basis (i.e., initially the TF of one user is increased and only when this user has completed the current transmission, the TF of the next user is increased).

3. System Model

The system simulation model considers two radio access bearers for supporting the interactive and conversational (real time) service classes, both with a maximum bit rate of 64 kbps in the uplink [7]. Representative applications for these two service classes are WWW browsing and videophone. The Transport Format Sets for both services are detailed in Table 1.

TABLE 1.

TRANSPORT FORMATS FOR THE CONSIDERED RABS.

Service		WWW	VIDEOPHONE
TrCH type		DCH	DCH
TB sizes, bit		336 (320 payload, 16 MAC/RLC header)	640
TFS	TF0, bits	0×336	0×640
	TF1, bits	1×336 (16 kb/s, SF=64)	2×640 (64 kb/s, SF=16)
	TF2, bits	2×336 (32 kb/s, SF=32)	-
	TF3, bits	3×336 (48 kb/s, SF=16)	-
	TF4, bits	4×336 (64 kb/s, SF=16)	-
TTI, ms		20	20

The interactive traffic model considers the generation of activity periods (i.e. pages for www browsing), where several information packets are generated, and a certain thinking time between them, reflecting the service interactivity. The specific parameters are: average thinking time between pages 30 s, average number of packet arrivals per page: 25, number of bytes per packet: average 366 bytes, maximum 6000 bytes (truncated Pareto distribution), time between packet arrivals: average 0.125 s, exponential distribution. The time between sessions is 300s. The conversational service consists in a constant bit rate source of 64 kbps with average duration 120s and exponential distribution. The conversational call rate per user is 30 calls/h, with Poisson arrivals. TB error rate target is 0.5% for both interactive and conversational services. The simulation model includes a cell with radii 0.5 km and intercell interference is represented by f=0.6. Physical layer performance, including the rate 1/3 turbo code effect, the 1500 Hz closed loop power control and a realistic channel impulse response estimator, is taken from [8]. The mobility model and propagation models are defined in [9], taking a mobile speed of 50 km/h and a standard deviation for shadowing fading of 10 dB. The maximum transmitted power of a mobile equipment is 21 dBm.

4. Results

The admission and congestion algorithms parameters considered in this Section are summarized in Table 2.

TABLE 2.

PARAMETERS CONSIDERED IN THE SIMULATIONS.

Admission control threshold, η_{max}	0.6
Congestion detection threshold η_{CD}	0.8
Congestion resolution threshold η_{CR}	0.7
ΔT_{CD}	10 frames
ΔT_{CR}	10 frames

In order to gain more insight into the differences between the two different admission algorithms, let consider the admission probability for an interactive or a conversational (real time) user. In the simulations, 5 real time users have been considered. As it can be seen in Fig. 1 and 2, the admission control without any kind of prioritization causes high number of rejections of conversational users when the load in the system is high.

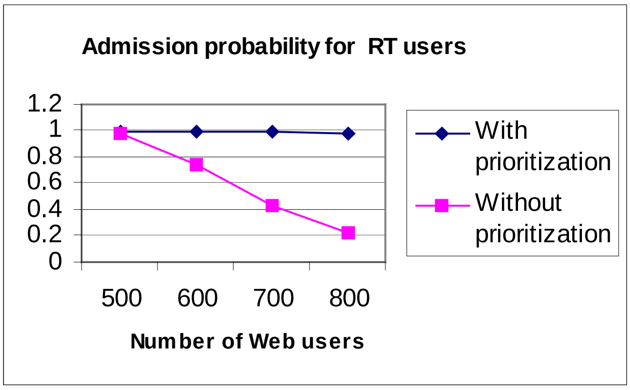

Fig. 1. Admission probability for conversational users.

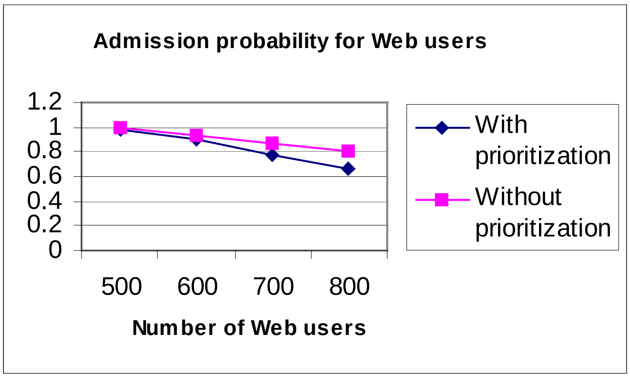

Fig. 2. Admission probability for interactive users.

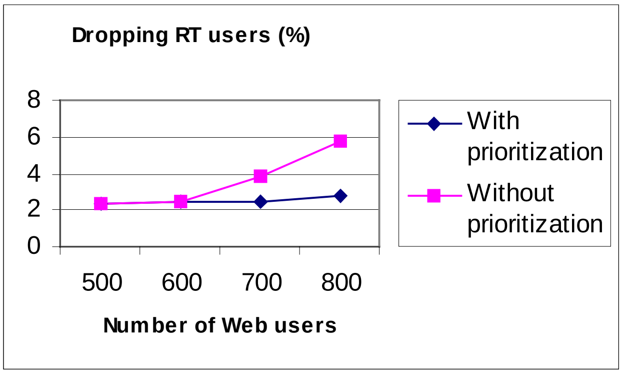

Fig. 3. Dropping for conversational users

However, the prioritization assures that the admission probability for conversational users is close to 100% by reducing the admission probability of interactive users. So, the prioritization algorithm takes resources from the interactive users in order to assure a predefined QoS to conversational users. It is worth mentioning that the system throughput with and without prioritization is the same, so that the admission procedure is able to get a capacity exchange from interactive to conversational users.

In addition to the admission probabilities, it is of interest to obtain some additional performance measurements of the system behaviour. In the case of conversational users, the dropping probability is observed: a conversational connection is dropped when the Eb/No obtained in this connection is Δ dB below the target Eb/No during Td consecutive frames. It has been considered Td=200 frames (2 seconds) and Δ=3dB. For interactive users, the average packet delay is considered as a representative performance figure.

According to Fig. 3, the dropping of conversational users is reduced if prioritization is considered in the admission algorithm. Notice that, with prioritization, the number of interactive users in the system is lower and the number of conversational users higher. Consequently, and since the interactive traffic sources are those that introduce a higher load variability, the load level in the system can be better controlled when prioritization exists, and a dropping reduction follows. Moreover, Fig. 4 shows a delay improvement for interactive users when prioritization is considered because congestion situations are less probable (see Fig. 5). The congestion probability is lower when prioritization is considered because, in this situation, the admission probability of interactive users is lower (see Fig. 2) and therefore the load in the system can be controlled in a better way.

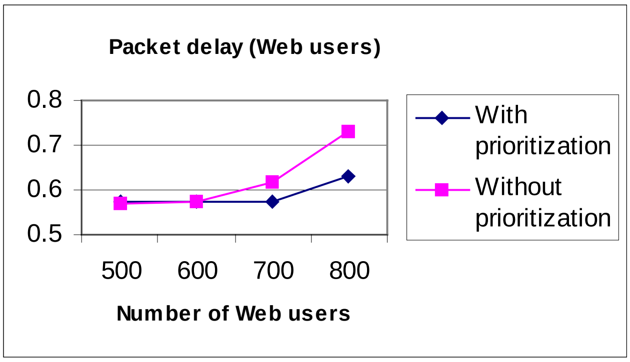

Fig. 4. Packet delay for interactive users.

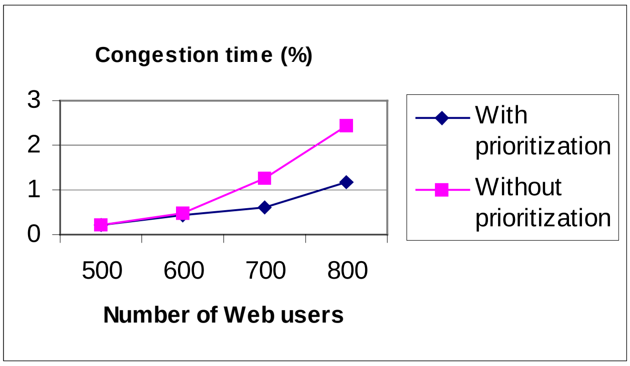

Fig. 5. Percentage of time in congestion resolution as a function of the number of interactive users

Although previous results have focused on admission control, the role played by the congestion control algorithm is also very important to explain the performance eventually obtained. In the following, two different scenarios will be studied under a prioritized admission control: one with the congestion control disabled and the other with the same congestion control than in the previous results. In both cases the admission scheme with prioritization is considered.

In Fig. 6, the dropping probability for 5 conversational users is shown as a function of the number of interactive users. If the number of interactive users is low, the congestion control does not improve the dropping probability, because this dropping is due to path loss instead of to excessive load in the system. However, when the number of interactive users and therefore the load variability is increased, the congestion control guarantees lower dropping probability for conversational users because the load in the system is controlled in a better way. So the congestion control mechanism is required to preserve the QoS of conversational service in a scenario with high load variability due to the existence of interactive users.

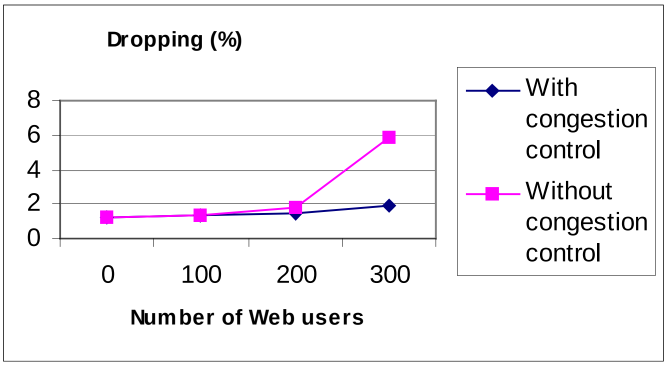

Fig. 6. Dropping of conversational users with and without congestion control

5. Conclusions

In the framework of Radio Resource Management strategies for W-CDMA systems, this paper has proposed a prioritization strategy in the admission control in order to assure resources to conversational users. It combines admission control with congestion control to make room for the conversational services at the expense of interactive services. This strategy provides a better system performance both in terms of admission and dropping of conversational services but even in terms of packet delay of interactive users by means of a reduction in the admission probability of these ones. Also, it has been shown that the existence of congestion control algorithms provides lower dropping to conversational users, so they are desirable to guarantee the QoS of real time services in the presence of traffic with high variability.

Acknowledgements

This work is part of the ARROWS project, partially funded by the European Commission under the IST framework (IST 2000-25133) and by the Spanish Research Council under grant TIC2001-2222.

References

[1] 3GPP TS 25.211, “Physical channels and mapping of transport channels onto physical channels (FDD)”

[2] 3GPP TR 25.922 v4.0.0, “Radio resource management strategies”

[3] O. Sallent, J. Pérez-Romero, F. Casadevall, R. Agustí, “An Emulator Framework for a New Radio Resource Management for QoS guaranteed Services in W-CDMA Systems”, *IEEE Journal on Selected Areas in Communications*, Vol.19, No. 10, October 2001, pp. 1893-1904.

[4] J. Pérez-Romero, O. Sallent, R. Agustí, “Admission Control for different UE-MAC Algorithms in UTRA-FDD”, 3rd International Conference on 3G Mobile Communication Technologies, London, May, 2002, pp. 256-260.

[5] H. Holma, A. Toskala (editors), *W-CDMA for UMTS*, John Wiley and Sons, 2000.

[6] V. Phan-Van, S. Glisic, “Radio Resource Management in CDMA Cellular Segments of Multimedia Wireless IP Networks”, WPMC 2001.

[7] 3G TS 34.108 v.3.2.0, “Common Test Environment for User Equipment. Conformance Testing”

[8] J. Olmos, S. Ruiz, “UTRA-FDD Link Level Simulator for the ARROWS Project”, IST’01 Conference Proceedings, pp. 782-787.

[9] 3GPP TR 25.942 v.2.1.3, “RF System Scenarios”