Performance Analysis of Radio Resource Management Strategies in UTRAN W-CDMA

J. Pérez-Romero, O. Sallent, R. Agustí, F. Casadevall
Universitat Politècnica de Catalunya
C/ Jordi Girona, 1 - 08034 Barcelona - Spain
E-mail : jorperez, sallent, ramon, ferranc @ tsc.upc.es

Abstract: **This paper focuses on the performance evaluation of Radio Resource Management (RRM) for W-CDMA. To this end, the framework to define RRM strategies is presented taking into account 3GPP specifications. The simulation model to characterise system performance in such an environment is also provided and finally some results are analysed for a selection of the presented strategies.**

Topics: RRM, QoS provisioning, Admission Control, Congestion Control, Packet Scheduling

I.-INTRODUCTION

The key feature of third generation mobile systems will be the ability to deliver wideband and high bit rate multimedia services alongside the traditional radio services such as voice, messaging and slow rate data. In that context, UMTS (Universal Mobile Telecommunication System) will provide wideband mobile multimedia services for the future mass market. The broad range of services expected to be supported through these 3G networks can be divided into four Quality of Service (QoS) classes: conversational (e.g. voice), streaming (e.g. video), interactive (e.g. www browsing) and background (e.g. e-mail). However, the provision of such mobile multimedia services under QoS guarantees will not be possible without a proper utilization of the air interface resources by means of Radio Resource Management (RRM) strategies that ensure the target QoS, the planned coverage area and offer a high system capacity [1][2]. Such strategies should deal with the peculiarities of the radio access technology, that in the UTRA FDD (UMTS Terrestrial Radio Access Frequency Division Duplex) mode of UMTS is based on W-CDMA (Wideband Code Division Multiple Access) [1]. One of the peculiarities of this access scheme is that it lacks from a constant value for the maximum available capacity, since it is tightly coupled to the amount of interference in the air interface. Therefore, RRM functions become crucial to manage this interference depending on the provided services. Moreover, RRM strategies are not subject of standardisation, so that they can be a differentiation issue among manufacturers and operators. Additionally, RRM functions can be implemented in many different ways, this having an impact on the overall system efficiency and on the operator infrastructure cost, so that definitively RRM strategies will play an important role in a mature UMTS scenario.

According to the 3GPP (3rd Generation Partnership Project) specifications, the radio interface of the UTRAN (UMTS Terrestrial Radio Access Network) is layered into three protocol layers: the Physical Layer (L1), the Data link Layer (L2) and the Network Layer (L3). Additionally, the layer 2 is split into two sub-layers, the Radio Link Control (RLC) and the Medium Access Control (MAC). On the other hand, the RLC and layer 3 protocols are partitioned in two planes, namely the User plane and the Control plane. In the Control plane, Layer 3 is partitioned into sublayers where only the lowest sublayer, denoted as Radio Resource Control (RRC), terminates in the UTRAN. Connections between RRC and MAC as well as RRC and L1 provide local inter-layer control services and allow the RRC to control the configuration of the lower layers. In the MAC layer, logical channels are mapped to transport channels. 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 instantaneous bit rates (i.e. different Spreading Factors SF) are associated to different TF. Since the User Equipment (UE) may have more than one transport channel simultaneously, whenever a combination of services is provided, the Transport Format Combination (TFC) refers to the selected combination of TF. The network assigns a list of allowed TFC to be used by the UE in what is referred as Transport Format Combination Set (TFCS) or Transport Format Set (TFS) if only one transport channel is considered. Note also that this set defines the maximum bit rate that can be used by the UE. Whenever a service is demanded by a certain UE, a Radio Access Bearer (RAB) should be allocated to it. The RAB defines the way how transmissions in the radio interface should be carried out in terms of type of transport channel, TTI, Transport Block size, possible TF (i.e. TFS), MAC and RLC headers, as well as all the physical layer aspects such as channel coding, interleaving, puncturing or slot formats.

Taking into account the constraints imposed by this radio interface architecture, the RRM functions are responsible of taking decisions regarding the setting of the parameters such as TF, TFS, etc. as well as other such as power level, code sequences, etc. RRM functions need to be consistent for both uplink and downlink, although the different nature of these links introduces some differences in the followed approach. In particular, RRM functions include:

- Admission control: It decides the admission or rejection of requests for setup and reconfiguration of radio bearers. The admission control procedure should take into account the impact of handover users and should be executed taking into account both uplink and downlink constraints.
- Congestion control: It faces situations in which the QoS guarantees are at risk due to the evolution of system dynamics (mobility aspects, increase in interference, traffic variability, etc.).
- Short term RRM mechanisms: They are devoted to decide the suitable radio transmission parameters for each connection in a reduced time scale and in a very dynamic way. Within these mechanisms the following functions can be included:
  - MAC algorithms: They are executed on a to decide the instantaneous Transport Format (or equivalently instantaneous bit rate) to be applied in each TTI for a given RAB.
  - Packet scheduling: It is responsible for scheduling non real time transmissions over shared channels. In UTRA FDD this functionality manages the occupation over the DSCH (Downlink Shared CHannel)
  - Power control: The purpose of this strategy is to optimise the mobile transmitted power (uplink) and the base station transmitted power (downlink). To this end, power control is executed in two steps [1]:
    - Inner loop power control: It is responsible of adjusting, on a fast time basis (i.e. each UTRA FDD 10 ms frame is subdivided into 15 slots each corresponding to a power control period), the transmitted power in order to reach the receiver with the required Eb/No target .
    - Outer loop power control: It is responsible of selecting a suitable Eb/No target depending on the BLER (BLock Error Rate) or BER (Bit Error Rate) requirement. It operates on a slower time basis than the inner loop power control, and adapts power control to changing environments.
- Code management: It is devoted to manage the downlink OVSF (Orthogonal Variable Spreading Factor) code tree used to allocate physical channel orthogonality among different users [1].
- Handover control: The purpose of this strategy is to optimise the cell or set of cells (i.e. the Active Set [1]) to which the mobile is connected.

The resulting decisions taken by RRM functions are executed by means of the radio bearer control procedures, which define the signaling messages to be exchanged between the network and the UE. Specifically, these messages are Radio Bearer Set-up, Physical Channel Reconfiguration and Transport Channel Reconfiguration [1]. Figure 1 summarizes the framework described above for the RRM definition.

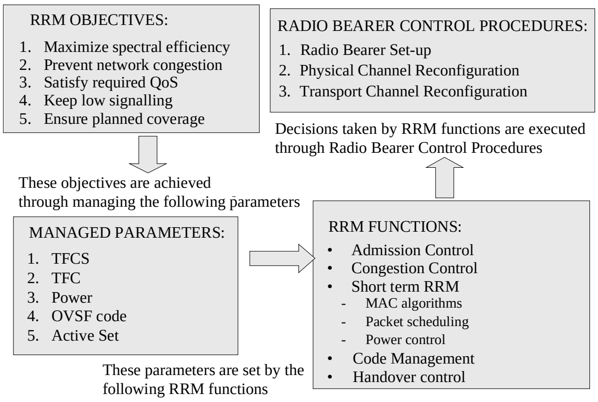


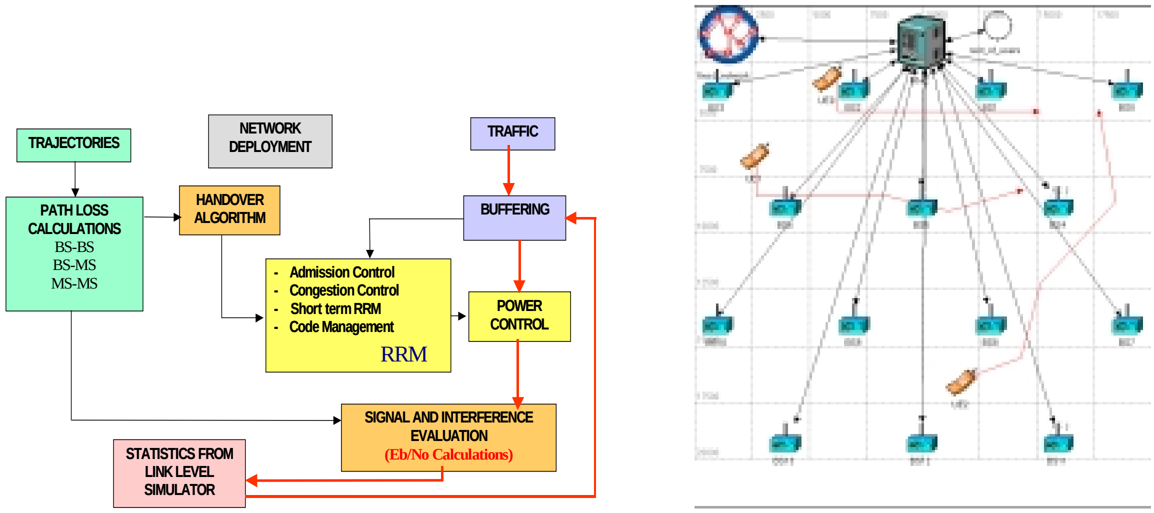
Figure 1. Framework for RRM definition

The complete characterisation of the achieved system performance is a complex problem with many mutual effects that depend on the specific solutions and parameters that have been considered for each of the RRM strategies. To this end, the evaluation of the whole system is usually split in two simulation tools: link and system level simulator. This paper mainly deals with the later, so that a multiuser, multicell and multiservice scenario is considered. In turns, the physical layer performance is modelled from the outputs of the link level simulator. The paper is organised as follows: section II presents the simulator that has been developed and section III provides some sample results dealing with admission control for both uplink and downlink, congestion control and packet scheduling over DSCH channel. Finally, section IV summarises conclusions.

II.- SIMULATION MODEL

The performance of the RRM algorithms has been evaluated through the use of a system level simulator that has been developed by means of the OPNET simulation platform. It allows the support of a wide range of RABs, traffic models as well as deployment scenarios. The simulator considers a representative service for each of the four defined traffic classes, according to specific RABs selected from TS 34.108 [1]. The mobility model and propagation models are defined in [3]. The physical layer characterisation is obtained through a link level simulator [4] that feeds the system level simulator with the transport block BLER statistics for each average (E_b/N_o). This characterisation includes a detailed simulation of all the processes involved at the physical layer, such as channel estimation, antenna diversity, rate 1/3 turbo coding as well as the 1500 Hz closed loop power control. Similarly, these link level results are also used to execute the outer loop power control in the system level simulator (i.e. to compute required E_b/N_o given a BLER requirement).

From a functional point of view, the procedures to be considered in the system level simulator are reflected in Figure 2(a). As an initial procedure, the network deployment module will be responsible for providing the simulator user with a mechanism to introduce the position of the different base stations and mobile stations as well as the other parameters characterizing the scenario to be evaluated. On the other hand, the RRM module is the core of the simulator responsible for executing the analyzed RRM strategies. The RRM module will act depending on the behavior of the mobile terminals in terms of traffic generation and mobility. Regarding the mobility issues, the simulator will contain modules to implement the trajectories of the terminals, to calculate the path loss to the base stations in the scenario and to decide the base stations in the active set depending on the handover algorithms. Similarly, traffic generation models will be simulated for each user depending on its corresponding service and the generated packets will be kept in buffers waiting for transmission. The RRM module will decide when and how the packets are transmitted through the radio interface. The power control mechanism will be responsible for determining the transmitted power of each transmission to reach a certain Eb/No target. Depending on this power and the position of the terminals the resulting Eb/No is evaluated for each transmission. Finally, the interaction with the off-line link level simulator results will decide the successful and erroneous transmissions. The buffers will be updated accordingly depending on the result for each transmission and on the availability of retransmissions.



(a) (b)

Figure 2. Functional Simulator Architecture (a) and Network Model (b)

Figure 2(b) shows a network model representing a possible scenario under test for the system level simulator. Particularly, the following nodes have been developed to implement the previously described functions: the RNC node, which simulates the behavior of the Radio Network Controller and particularly deals with the simulator control and with the RRM strategies, the BS node, the UE node, which simulates the behavior of a User Equipment in terms of mobility, traffic generation and radio transmission functions, and the Fixed Network node, which acts as the generation source for downlink traffic. Furthermore, the "Rest of Users" node simulates a high number of UEs without having to locate all of them in the network model. The only functional difference between a UE node and a user in the Rest of Users node is that the first follows a user-defined trajectory while the second makes use of a mobility model.

III.- PERFORMANCE RESULTS

This section provides an overview of some of the results that have been obtained with the previously described simulation tool for the different RRM strategies. Since RRM for W-CDMA is a complex problem with many mixing effects, this section is organised as a collection of different case studies each one focusing on a given RRM function either in uplink or in downlink. This will allow to devise the main parameters and factors influencing on the achieved performance as well as the key elements of the different RRM strategies. The results presented here consider an UL/DL 64 kb/s CBR videophone service (representative of the conversational service class) and a WWW service (representative of the interactive class) with a maximum bit rate of 64 kb/s in the UL and 256 kb/s in the DL. Simulation parameters are summarised in Table 1.

Table 1. Simulation parameters

Scenario size		2.25 km x 2.25 km	
Chip Rate W	3.84 Mcps	**Frame duration**	10 ms
BS parameters		**UE parameters**	
Cell radius	500 m	Maximum transmitted power	21 dBm
Cell type	Omnidirectional	Minimum transmitted power	-44 dBm
Maximum transmitted power	43 dBm	Thermal noise	-100 dBm
Thermal noise	-106 dBm	Mobile speed	50 km/h
Common Control Channels Power	30 dBm	DL Orthogonality factor	0.4
Propagation model (macrocellular)			
Shadowing deviation	10 dB	Shadowing decorrelation length	20 m
Handover parameters (conversational)		**Handover parameters (interactive)**	
Active Set (AS) maximum size	2	Active Set maximum size	1
Threshold to enter AS	3 dB	Replacement hysteresis	1 dB
Hysteresis to enter AS	1 dB	Time to Trigger Handover	0.5s
Replacement hysteresis	1 dB		
Time to Trigger Handover	0.5s		
Traffic model (conversational)		**Traffic model (interactive)** [3]	
Call duration	120s (exponential)	Number of pages per session	5 (geometrical)
Offered bit rate	64 kb/s (CBR)	Reading time between pages	30 s (exponential)
Activity factor	1	Number of packets per page	25 (geometrical)
Call rate	29 calls/h/user (Poisson arrivals)	Time between packets (exponential)	0.0625 s (DL) 0.125 s (UL)
		Packet length	Pareto distributed (α=1.1, k=81.5 bytes, m=6000 bytes)
		Session rate	25 sessions/h/user
QoS parameters (conversational)		**QoS parameters (interactive)**	
BLER target	1%	Average page delay	<4s
Dropping probability	<1%	Dropping probability	<1%
Dropping condition	1s below Eb/No target or lack of codes in the new cell during HO	Dropping condition	Lack of codes in the new cell during HO

III.1.- Uplink admission control

In W-CDMA there is a close relationship between coverage and capacity, since in such an interference limited scenario, the power required by a given connection is directly related to the number of users that are simultaneously transmitting. Specifically, the power to be transmitted by each UE of a given cell depends on the cell uplink load factor η_{UL} and the path loss $L_p(d_i)$ (including shadowing) at distance d_i as [5]:

$$P_{T,i} = L_p(d_i)\frac{P_N \dfrac{1}{1-\eta_{UL}}}{\dfrac{W}{\left(\dfrac{E_b}{N_o}\right)_i R_{b,i}}+1} \qquad i=1..n \qquad (1)$$

where n is the number of users transmitting simultaneously at a given moment, $R_{b,i}$ is the i-th user instantaneous bit rate, W is the total bandwidth after spreading and P_N is the thermal noise power. The uplink load factor provides the theoretical spectral efficiency of a W-CDMA cell and is given by [7]:

$$\eta_{UL} = \left(1+\frac{\chi}{P_R}\right)\sum_{i=1}^{n}\frac{1}{\dfrac{W}{\left(\dfrac{E_b}{N_o}\right)_i R_{b,i}}+1} = \frac{P_R+\chi}{P_R+\chi+P_N} \qquad (2)$$

where χ is the intercell (other-cell) interference and $(E_b/N_o)_i$ stands for the i-th user requirement. P_R is the total received own-cell power at the base station

As it is observed in equation (1), capacity and coverage are closely related in W-CDMA networks, and therefore both must be considered simultaneously. The coverage problem is directly related to the power availability, so that the power demands deriving from the system load level should be in accordance with the planned coverage. So, it must be satisfied that the required transmitted power will be lower than the maximum power available at the UE and high enough to be able to get the required (E_b/N_o) target even at the cell edge. This can be achieved by means of a proper control of the cell load factor, by means of admission and congestion control strategies.

From the implementation point of view, admission control policies can be divided into modelling or statistical-based and measurement-based policies [6]. In both approaches, uplink admission control is typically based on cell load factor monitoring, so assuming that K users are already admitted in the system, the admission control algorithm considers the increment in the load factor that the new acceptance would originate. Then, the condition to be checked for the admittance of the $(K+1)$th request would be:

$$\eta_{UL} + \Delta\eta \le \eta_{\max} \qquad (3)$$

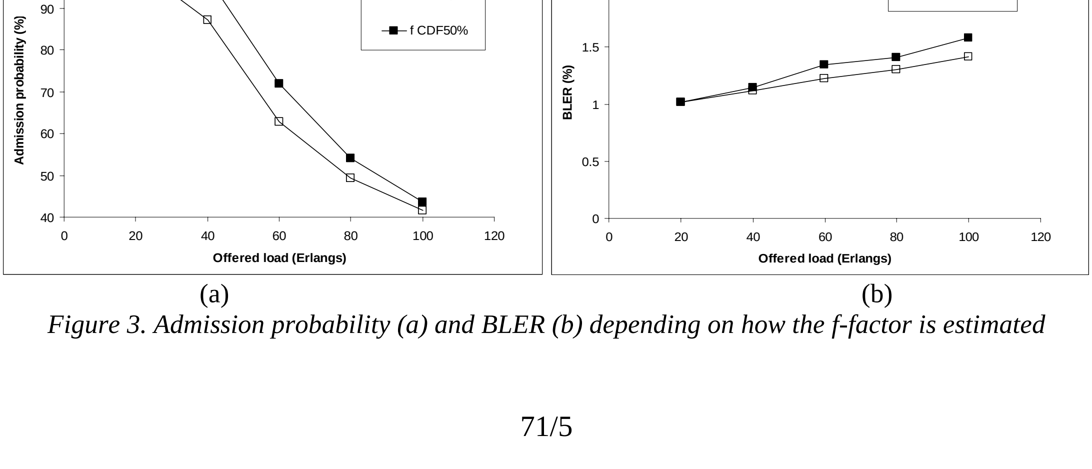
η_{UL} being the current estimation of the uplink cell load factor, $\Delta\eta$ being the estimated contribution demanded by the $(K+1)$th user and $\eta_{\max}$ the admission control threshold.

Focusing on the statistical-based approach and in a single service scenario, the load factor for the K users that are currently connected to the cell can be statistically estimated by characterising every connection with a certain activity factor, v_i, and a certain transmission rate, $R_{b,i}$. The intercell to intracell interference contribution is characterised by the so-called f-factor ($f=\chi/P_R$), considered for example in [8]. The estimation of η_{UL} and the contribution $\Delta\eta$ demanded by the $(K+1)$th user are then given by:

$$\eta_{UL} = (1+f)\sum_{i=1}^{K} \frac{1}{\dfrac{W}{v_i \cdot \left(\dfrac{E_b}{N_o}\right)_i R_{b,i}} + 1} \qquad (4)$$

$$\Delta\eta = \frac{1+f}{\dfrac{W}{v_{K+1} \cdot \left(\dfrac{E_b}{N_o}\right)_{K+1} R_{b,i}} + 1} \qquad (5)$$

It should be strengthened here that the estimated cell load, η_{UL}, that is used for admission purposes, will be different from the real cell load, not only instantaneously but even in average terms, if the admission control parameters (such as f, $R_{b,i}$, etc.) are not suitably adjusted. For example, in the case that only videophone service is considered, the main parameter that has an influence on the statistical admission control is the other-to-own cell interference factor f, since the service is of constant bit rate nature and the traffic source is always active along the connection time. Because of the system dynamics as well as the handover procedures, the parameter f exhibits large variations. In this case, the average value is not representative enough and it can provide rather pessimistic results in terms of admission. It is shown that a better setting would be to make use of the 50% percentile (i.e., 50% of the Cumulative Distribution Function). The improvement obtained with this method is depicted in Figure 3 (a) and (b) where simulations show that the average f value highly degrades the admission probability without a significant improvement in terms of BLER, specially for medium loads. On the contrary, with the 50% percentile the admission probability is improved with a negligible BLER degradation.



(a) (b)

Figure 3. Admission probability (a) and BLER (b) depending on how the f-factor is estimated

III.2.- Downlink admission control

Despite some uplink concepts can be applied to downlink, significant differences arise. The restrictions imposed by each link are not of the same nature: while in the downlink the maximum transmitted power is the same regardless the number of users, in the uplink each user has its own power amplifier. Therefore, as the transmitted power should be shared in the downlink among all the users, their instantaneous locations have a high impact over the performance of the rest of users, even for low loads, while in the uplink a particular user location has only impact over its own performance. As a result, the cell load as well as the capacity demanded by a given user varies as this user moves around the cell and, consequently, the user location influences on the amount of radio resources that should be allocated to it. More specifically, for *n* users receiving simultaneously in a given cell, the following inequality for the i-th user must be satisfied [5]:

$$\frac{\dfrac{P_{Ti}}{L_p(d_i)}\dfrac{W}{R_{b,i}}}{P_N+\chi_i+\rho\times\left[\dfrac{P_T-P_{Ti}}{L_p(d_i)}\right]}\geq\left(\frac{E_b}{N_o}\right)_i \qquad (6)$$

P_T being the base station transmitted power including the power P_c devoted to pilot and common control channels, P_{Ti} being the power devoted to the i-th user, χ_i representing the intercell interference observed by the i-th user, $L_p(d_i)$ being the path loss at distance d_i (including shadowing) and P_N the background noise. ρ is the orthogonality factor since orthogonal codes are used in the downlink direction. Notice that, differently from the uplink case, in downlink the intercell interference is user-specific. Additionally, physical limitations into the power levels are given by the maximum base station transmitted power, P_{Tmax}. Then, it can be obtained that the total transmitted power to satisfy all the users demands should be:

$$P_T=\frac{P_c+P_N X}{(1-\eta_{DL})} \qquad (7)$$

where the downlink load factor η_{DL} and the factor X that depends on user locations are defined as:

$$\eta_{DL}=\sum_{i=1}^{n}\frac{\left(\rho+\dfrac{\chi_i\times L_P(d_i)}{P_T}\right)}{\dfrac{W}{\left(\dfrac{E_b}{N_o}\right)_i R_{b,i}}+\rho} \qquad (8)$$

$$X=\sum_{i=1}^{n}\frac{L_p(d_i)}{\dfrac{W}{\left(\dfrac{E_b}{N_o}\right)_i R_{b,i}}+\rho} \qquad (9)$$

It can be observed that as the load factor increases the power demands also increase. On the other hand, for a maximum transmitted power, $P_{T,max}$, eq. (7) provides the maximum allowable cell load factor, which shows a quite sharp variation depending on *X*. Therefore, the maximum cell load factor and the transmitted power are coupled in the downlink and related by a time varying factor that depends on user locations. Consequently, it may seem more reasonable to control the downlink operation through the transmitted power rather than through the cell load factor, as it uses to be the case in the uplink. Within this context, the considered admission control algorithm checks the following condition to decide the acceptance of a new connection request in the system, at the i-th frame:

$$P_{AV}(i)+\Delta P_T(i)\leq P_T^*(i) \qquad (10)$$

where $P_{AV}(i)$ is the average transmitted power during the last *T* frames, $\Delta P_T(i)$ is the power increase estimation due to the new request (notice that it may vary along time) and $P_T^*(i)$ is the admission threshold.

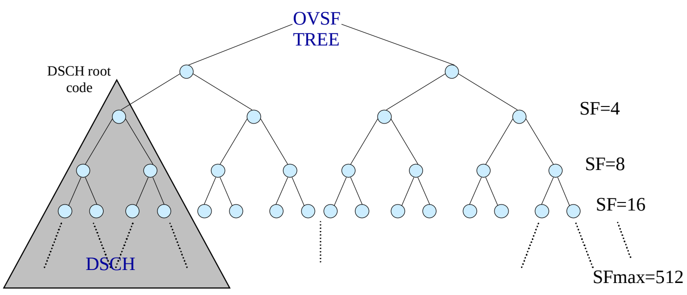
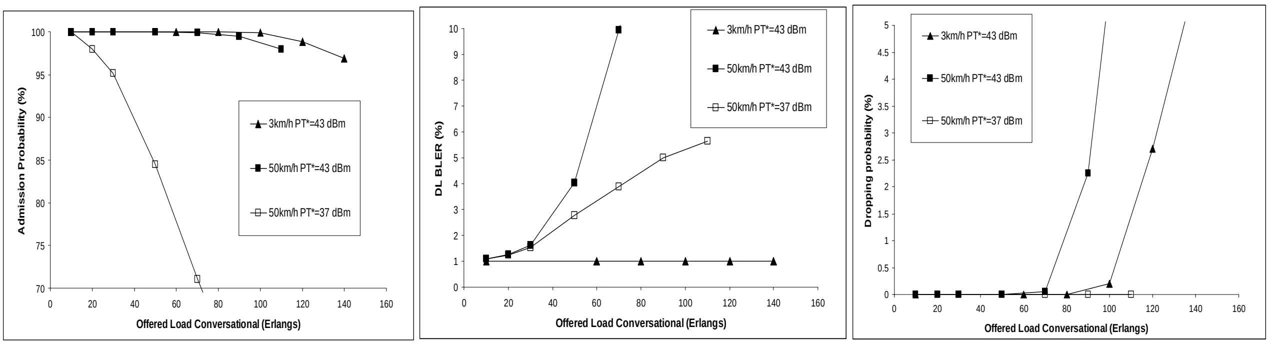


Figure 4. OVSF code tree

In the downlink direction of UTRA FDD simultaneous transmissions are distinguished by means of different OVSF codes, which are generated according to a tree structure as depicted in Figure 4. So, in addition to power availability given by equation (10), code availability must also be considered for admission control purposes. The OVSF code tree has the property that two or more codes belonging to different tree branches are orthogonal, while codes belonging to the same branch do not keep orthogonality. The higher the spreading factor, the higher the number of available codes in the tree (so there can be 4 orthogonal codes with SF=4, 8 with SF=8, and so on until reaching the maximum spreading factor SF_{max}=512). When mapping transport channels onto OVSF codes, part of the tree will be devoted to common channels (e.g. the pilot or the broadcast channel, each of them requiring a code with SF=256) and the remainder to shared and dedicated channels (DCH). Services of real time nature (e.g. conversational) are mapped to dedicated channels, thus occupying a part of the code tree, while services of non real time nature (e.g. interactive) may take advantage of the availability of the DSCH, which occupies another part of the tree, as depicted in Figure 4, starting from the DSCH root code whose spreading factor is denoted as SF_{root} and it is fixed depending on the specific needs in terms of the provided services. Transmissions of over the DSCH are time and code multiplexed and the codes are assigned dynamically by a packet scheduling algorithm that takes into account the QoS that each flow is expecting. In any case, all the users that transmit in the DSCH must have also a low bit rate DCH allocated in order to carry control information, so that interactive users will also take some part of the code tree out of the DSCH branches. DSCH management will be analyzed in Section III.4.



(a) (b) (c)

Figure 5. Performance results when only conversational users are considered

Furthermore, depending on the considered environment, either OVSF codes or power availability may be the limiting factor. As an example Figure 5 depict performance results in a DL scenario with only conversational traffic, so that no DSCH is considered, for two different situations: with mobile speed 3 km/h and 50 km/h. The period to average power measurements for admission control is 1s. Results are given in terms of admission probability, BLER and dropping probability. It can be observed that for the 3 km/h case, no BS power limitations occur which leads to always ensuring the required BLER of 1% even for high loads. However, for such high loads, dropping probability increases (and admission decreases) because of the lack of OVSF codes in the new cell during a handover process. On the contrary, if the mobile speed is 50 km/h, performance degrades for lower loads due to power limitations. The reason is twofold: first of all, the required Eb/No is higher for 50 km/h, and, additionally, and due to the delay in the handover process, a fast user that enters a new cell is highly interfered while waiting for the addition of the cell to the Active Set. As a result, it is important to properly select the admission threshold P_T* in order to reduce the BLER at the expense of a reduced admission probability. For example, see in Figure 5(b) the improvement in terms of BLER when considering a P_T*=37 dBm. Notice also in that in this case, due to a reduced admission probability OVSF occupation is also lower and therefore droppings do not occur (see Figure 5(c)).

III.3.- Uplink Congestion Control

Congestion control mechanisms should be devised to face situations in which the system has reached an overload status and therefore the QoS guarantees are at risk due to the evolution of system dynamics. The congestion state then has to invoke a congestion control procedure, which should set the system operation point again at a stable situation. More specifically, the following parts are identified:

1. Congestion detection: A certain criterion is introduced in order to decide that the network is in congestion and as a result to trigger a congestion resolution procedure. Specifically, it is assumed that the system has entered the congestion situation when the load factor increases over a certain threshold during a certain amount of time, i.e. if $\eta \geq \eta_{CD}$ during a certain percentage *p* of the frames within ΔT_{CD}. Usually the

network is planned to operate below a certain maximum load factor and the congestion detection threshold, η_{CD}, should be set in accordance to this maximum planned value.

2. Congestion resolution: When a congestion is assumed in the network, the congestion resolution algorithm executes a set of rules to lead the system out of the congestion status, a part from blocking all new requests. Three steps are identified:

i. Prioritisation: Ordering the different users from lower to higher priority (i.e., from those that expect a lower grade of service to those with more stringent QoS requirements) in a prioritisation table.

ii. Load reduction: This is achieved by means of a reduction of the maximum transmission rate (i.e. limiting the TFS) capabilities of delay-tolerant users already accepted in the network, beginning from the top of the prioritisation table. The considered load reduction algorithm first reduces completely the TFS of a user before going to the next user in the table. This reduction is executed until the estimated load factor is below a given threshold η_{CR}.

iii. Load check: After the previous actions, one would check again the conditions that triggered the congestion status. If congestion persists, one would continue with the next user in the prioritisation table. It is considered that the overload situation has been overcome if during a certain percentage *p* of the frames within ΔT_{CR} the load factor is below the threshold η_{CR}.

3. Congestion recovery: Once the congestion resolution phase decides that the congestion situation has been overcome, a congestion recovery algorithm is needed to restore to the different mobiles the transmission capabilities they had before the congestion was triggered. Such an algorithm is crucial because depending on how the recovery is carried out the system could fall again in congestion. The considered congestion recovery algorithm can be regarded as a “time scheduling” algorithm, restoring the transmission capabilities on a user by user approach. That is, initially a specific user is again allowed to transmit at maximum rate. Once this user has emptied the transmission buffer, another user is allowed to recover the maximum rate and so on.

In the following some results to attain the key congestion control parameters to be taken into account are presented. The considered scenario includes a mix of conversational and interactive traffic, in order to reflect the impact that a variable bit rate source has over conversational traffic. Particularly, it should be pointed out that interactive traffic has some room to execute retransmissions in case that packets are erroneously received (e.g. if load factor is very high). On the contrary, conversational traffic, with stringent delay requirements, has no room for retransmissions and therefore high load factor values due to traffic variability may seriously degrade the achieved performance. With the previously described congestion control algorithm, different policies are compared with different thresholds for triggering and releasing the congestion status (i.e. low values like η_{CD}=0.75 and η_{CR} =0.6, or high values like η_{CD}=0.9 and η_{CR} =0.75) and with different time periods before triggering the congestion recovery phase (i.e. low values like ΔT_{CR} =0.1s and ΔT_{CR} =1s or high values like ΔT_{CR} =10s). In all the cases, it has been assumed ΔT_{CD} =0.1s and a percentage of time *p*=90% to trigger the different events, which from previous simulations (not shown for the sake of brevity) reveal to be suitable values. The results for the different possibilities are shown in Table 2.

In the no congestion control case, all conversational and interactive requests are accepted. Then, if the system dynamic evolves freely, the conversational BLER increases significantly beyond the target value and the same can be said for the interactive BLER (see Table 2). However, the impact on interactive users is negligible because the retransmissions only increase the average packet delay very slightly. When congestion control policies are adopted, the effects are a BLER reduction for conversational users, an average packet delay increase for interactive users (because congestion control reduces the bit rate of interactive users) and, since requests are blocked during congestion periods, a reduction of the admission probability.

With respect to the congestion resolution period ΔT_{CR}, a safe period of ΔT_{CR}=10s severely penalizes the admission rate of both interactive and conversational users. Note that a high value for ΔT_{CR} makes that system-declared congestion situations last longer. Also, a safe period of ΔT_{CR}=10s severely penalizes the average packet delay of interactive traffic, because during longer periods interactive users have limited maximum bit rate capabilities. On the contrary, this safe period is able to keep the conversational BLER closer to its target value. A short period like ΔT_{CR}=0.1s provides higher admission rates and a lower average interactive packet delay at the expense of a higher conversational BLER, which may raise up to 1.31% in some of the analyzed cases. This BLER increase arises because the congestion situations are not so well controlled (it can be decided after ΔT_{CR}=0.1s that the congestion has been overcome although in a short period the algorithm will likely trigger again congestion).

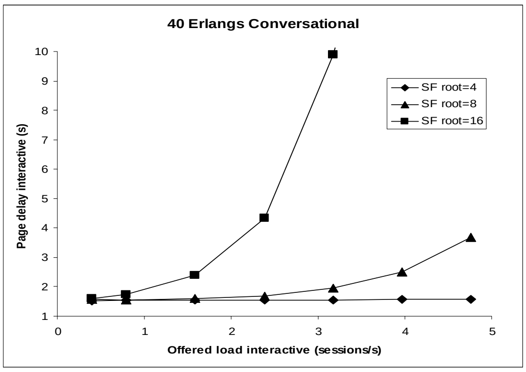
With respect to the congestion cell load triggering thresholds η_{CD} and η_{CR}, it can be said that since η_{CD}=0.9 and η_{CR}=0.75 constitute a late congestion trigger compared to η_{CD}=0.75 and η_{CR}=0.6, the conversational BLER degrades more in the former case, BLER=1.31%, than in the later BLER=1.15% (i.e. when the system triggers congestion, the cell load has already remained at high values for a certain period of time and this has caused some erroneous transmissions). On the other hand, this late detection avoids some users to be blocked and, consequently, in the former case a higher admission probability is found. In turns, for the late detection case of η_{CD}=0.9 and η_{CR}=0.75 a lower interactive average packet delay is obtained. This is because interactive delay is more degraded because of the congestion control actions (i.e. restricted transmission capabilities) than because of packet retransmissions due to too much load. Anyway, a better control of the conversational transmissions (i.e. lower BLER) is achieved.

Table 2. Performance figures for different congestion control policies and parameters.

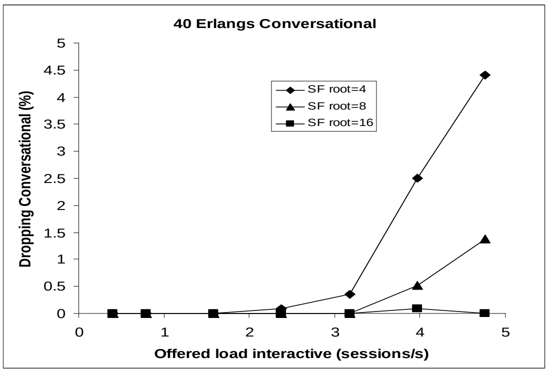
3.5 sessions/s interactive 20 Erlangs conversational		**Admission Conv. (%)**	**Admission Interact. (%)**	**BLER Conv. (%)**	**BLER Interact. (%)**	**Delay Interact. (s)**
No Congestion control		100	100	2.40	5.67	0.14
η_{CD}=0.75 η_{CR} =0.6	ΔT_{CR}=0.1s	93	57	1.14	1.47	0.47
	ΔT_{CR}=1s	69	48	1.08	1.30	1.12
	ΔT_{CR}=10s	42	37	1.02	1.14	3.14
η_{CD} =0.9 η_{CR}=0.75	ΔT_{CR}=0.1s	96	68	1.31	2.00	0.29
	ΔT_{CR}=1s	80	58	1.17	1.67	0.66
	ΔT_{CR}=10s	52	49	1.09	1.33	1.83

III.4.- Packet Scheduling in the DSCH channel

As explained in section III.2, OVSF code management plays an important role in the admission process. Particularly, the DSCH channel comprises a part of the OVSF code tree that is reserved for non real time users (e.g. interactive) and therefore it cannot be allocated to DCH channels. Transmissions in the DSCH channel are determined by means of a packet scheduling algorithm that takes into account not only code but also power availability. It is worth mentioning that when mixing real time and non real time traffic a trade-off arises in the suitable allocation of the DSCH channel given by the spreading factor of the DSCH root code SF_{root} (see OVSF code tree in Figure 4). When SF_{root} is low, there is more room to allocate interactive transmissions, while on the other hand there are less resources to be allocated to DCH channels for conversational users. This lack of DCH channels may originate rejections of new connections or even call droppings when a user must handoff a call to a cell where there are not available codes.



(a)



(b)

Figure 6. System performance for different DSCH configurations

To illustrate this situation, Figure 6 shows the performance results for conversational and interactive traffic for different interactive offered loads with 40 Erlangs of conversational traffic. Different values for the SF_{root} of the DSCH channel have been considered. P_T^*=43 dBm has been assumed in the admission control. Mobile speed is 3km/h. The scheduling algorithm is described in [9]. From the point of view of interactive traffic, their delay increases when SF_{root} is high, since the scheduler has less room to accommodate transmissions in the DSCH. However, depending on the selected SF_{root}, the system may run out of codes and the admission begins to decrease and droppings may occur due to lack of OVSF codes in the new cell during a handover process (see Figure 6(b)). It is worth mentioning that dropping affects both to conversational and interactive traffic since for the latter also a low bit rate DCH control channel with SF=256 must exist. In any case, in this scenario the system is not limited by power availability since BLER equals the target value of 1%, meaning that the power transmitted by the base station is enough to keep all users requirements. Therefore, the key parameter influencing performance becomes SF_{root} rather than P_T^*. So in a mixed service scenario different regions of usage of each SF_{root} can be obtained depending on the total offered load of interactive and conversational traffic in order to ensure some quality criterion. Specifically, Figure 7 depicts these regions for a dropping limit of 1% and an average page delay lower than 4s. Notice that for low conversational loads, it is the SF_{root}=4 that maximizes performance since it ensures the required average page delay. On the contrary, when conversational load increases, higher SF_{root} values are more suitable in order to avoid call droppings.

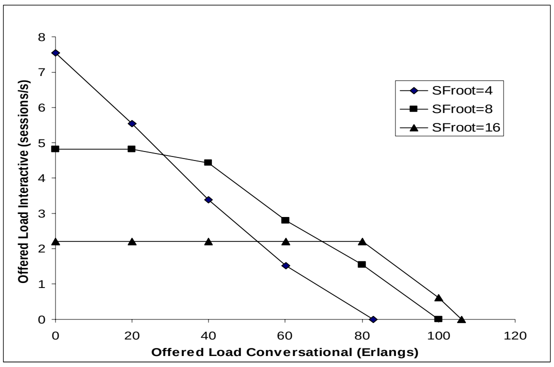


Figure 7. Regions for the selection of the DSCH SF_{root} depending on the traffic mix

IV.- CONCLUSIONS

This paper has presented the framework considered in 3GPP for the definition of RRM strategies that ensure QoS guarantees in UTRAN W-CDMA. A simulation environment for the performance evaluation of such strategies has been presented together with some representative results. With respect to uplink admission control an appropriate setting of a statistical load factor estimation method has been given, while for downlink admission control different situations in which the limiting factor are either the OVSF codes or the power availability have been analysed. Uplink congestion control has also been studied, leading to the conclusion that it is critical to ensure BLER requirements in a mixed traffic scenario. Finally, DSCH channel configuration has been analysed, resulting in different optimal configurations depending on the traffic mix.

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 CICYT (TIC2001-2222).

REFERENCES

[1] 3GPP specifications. http://www.3gpp.org

[2] O. Sallent, J. Pérez-Romero, R. Agustí, F.Casadevall, "Provisioning Multimedia Wireless Networks for Better QoS: RRM Strategies for 3G W-CDMA", *IEEE Communications Magazine*, February 2003.

[3] UMTS 30.03 v3.2.0 TR 101 112 "Selection procedures for the choice of radio transmission technologies of the UMTS", ETSI, April, 1998.

[4] J.J. Olmos, S. Ruiz, "Transport block error rates for UTRA-FDD downlink with transmission diversity and turbo coding", *13th IEEE PIMRC Conference*, Vol. 1 , 2002, pp. 31-35.

[5] 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.

[6] V. Phan-Van, S. Glisic, "Radio Resource Management in CDMA Cellular Segments of Multimedia Wireless IP Networks", *WPMC'01 Conference Proceedings*

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

[8] A.J. Viterbi, A. M. Viterbi, E. Zehavi, "Other-Cell Interference in Cellular Power-Controlled CDMA", *IEEE Transactions on Communications*, Vol. 42, No. 2/3/4, February/March/April, 1994, pp. 1501-1504.

[9] J. Pérez-Romero, O. Sallent, R. Agustí, "Downlink Packet Scheduling for a Two-Layered Streaming Video Service in UMTS", IST Mobile & Wireless Telecommunications Summit 2002 (IST 2002), Thessaloniki, Greece, June 2002, pp. 212-216.