AVERAGE AND PEAK INTERFERENCE MANAGEMENT IN W-CDMA UMTS

J. Pérez-Romero, O. Sallent, R. Agustí, J. Sánchez

Universitat Politècnica de Catalunya
c/ Jordi Girona 1-3, 08034 Barcelona, Spain
email: [jorperez, sallent, ramon @ tsc.upc.es]

Abstract

The definition and assessment of suitable Radio Resource Management (RRM) strategies able to provide a required 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 particular, the admission control of a new connection, the congestion control mechanisms to deal with peak interference situations and the dynamic management of the transmission parameters are studied by considering as a reference interactive-like 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 management of the transmission parameters (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 QoS, not much effort has been devoted to them up to date in the open literature, specially when all of them are jointly considered. Within this context, this paper proposes new admission and congestion control mechanisms whose effects are studied in an uplink scenario that also includes proposed decentralised UE-MAC strategies. The paper is organised 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 summarises the results obtained.

2.- RRM ALGORITHMS

UMTS provides a layered architecture where logical channels are mapped to transport channels in the MAC layer. A transport channel defines the way in which 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 a 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.

Focusing in the uplink direction, centralized solutions (i.e. RRM algorithms located at the RNC) may provide better performance compared to a distributed solution (i.e. RRM algorithms located at the UE) because much more RRM relevant information related to all users involved in the process may be available at the RNC. In return, executing decisions taken by RRM algorithms would be much more costly in terms of control signaling because in this case UE must be informed about how to operate. Consequently, strategies face with the performance/complexity trade-off, which usually finds a good solution in an intermediate state where both centralized and decentralized components are present. 3GPP approach for the uplink could be included in this category, as it can be divided in two parts:

1. Centralized component (located at RNC). Admission and congestion control are carried out.
2. Decentralized part (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.

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 and the estimate of the load increase that the establishment of the bearer request would cause in the radio network [4]. From the implementation point of view, admission control policies can be divided into modeling-based and measurement-based policies [5].

In case the air interface load factor η is estimated in statistical terms and assuming that K users are already admitted in the system, the (K+1)th request should verify:

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

where other-cell interference power is modeled as a fraction *f* of the own-cell received power, *(E_b/N_o)* is the target quality level and *r* the coding rate. According to (1) different admission strategies arise by balancing the following parameters: a) the spreading factor: by setting SF_i as an estimated average value the user will adopt along its connection time the assumed load will be closer to the real situation at the expense of relying on the statistical traffic multiplexing. In turns, considering SF_i as the lowest SF in the defined RAB covers the worst case at the expense of overestimating the impact of every individual user and, consequently, reducing the capacity, b) the activity factor of the traffic source: by setting v_i <1 the admission procedure can be closer to the real situation of discontinuous activity (typical in interactive-like services) at the expense of relying on the statistical traffic multiplexing. In turns, v_i =1 covers the worst case at the expense of overestimating the impact of every individual user and, consequently, reducing the capacity, c) the overall load level: by setting η_{max} the admission procedure allows for some protection against traffic multiplexing situations above the average (for example having more active connections than the expected average number, or having more users making use of low SF than the expected number).

2.2.- Congestion control

Congestion control mechanisms should be devised to face peak interference situations in which the system has reached a congestion status and therefore the QoS guarantees are at risk due to the evolution of system dynamics (mobility aspects, increase in interference, etc.). Congestion occurs when the admitted users can not be satisfied with the normal agreed services for a given percentage of time because of an overload. The congestion state then has to invoke some procedure that could prevent some users from getting the normal QoS margin not beyond the contracted percentage of time. The overload causing congestion can be due to several facts, for example: a) in the admission procedure the declared values (*SF_i*, *v_i*) have not been precisely specified, so this may cause an interference level higher than the accounted for in the admission phase or b) an intercell interference increase. 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 criteria 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 during a certain amount of time, *ΔT_{CD}*: $\eta \ge \eta_{CD}$.

2) Congestion resolution. When a 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) 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 prioritization table.

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.

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 could be 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: $\eta \le \eta_{CR}$.

3) Congestion recovery: A congestion recovery algorithm is needed in order to restore to the different mobiles the transmission capabilities 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.

2.3.- UE-MAC strategy

For the decentralized uplink RRM component, few studies aligned to 3GPP specifications are available in the open literature. For interactive-like services, a specific algorithm referred as Service credit (SCr) algorithm is proposed: when a certain mean bit rate should be guaranteed, a new possibility arises that makes use of the “service credit” (SCr) concept. The SCr of a connection accounts for the difference (measured in TB per TTI) between the obtained bit rate and the expected bit rate for this connection. Essentially, if SCr < 0 the connection has obtained a higher bit rate than expected, if SCr > 0 the connection has obtained a lower bit rate than expected. At the beginning of the connection: SCr(0)=0. In TTI=n, the *SCr(n)* for a connection should be updated as follows:

SCr(n) = SCr(n-1) + (Guaranteed_rate/TB_size) - Transmitted_TB(n-1) (2) where *Guaranteed_rate* is the number of bits per TTI that would be transmitted at the guaranteed bit rate, *TB_size* is the number of bits of the Transport Block for the considered RAB and *Transmitted_TB(n-1)* is the number of successfully transmitted Transport Blocks in the previous TTI. As a result, SCr(n) is a measure of the number of Transport Blocks that the connection should transmit in the current TTI to keep the guaranteed bit rate. For example, if TB_size=240 bits, Guaranteed_rate=24 Kb/s, and TTI=20 ms, the UE adds 2 service credits each TTI.

Then, assume that in the buffer there are L_b bits, the number of Transport Blocks to be transmitted in the current TTI=n would be:

$$numTB = min\left(\left\lceil \frac{L_b}{TBsize} \right\rceil, SCr(n), TBmax\right) \quad (3)$$

Once *numTB* is calculated, the determination of TF is straightforward.

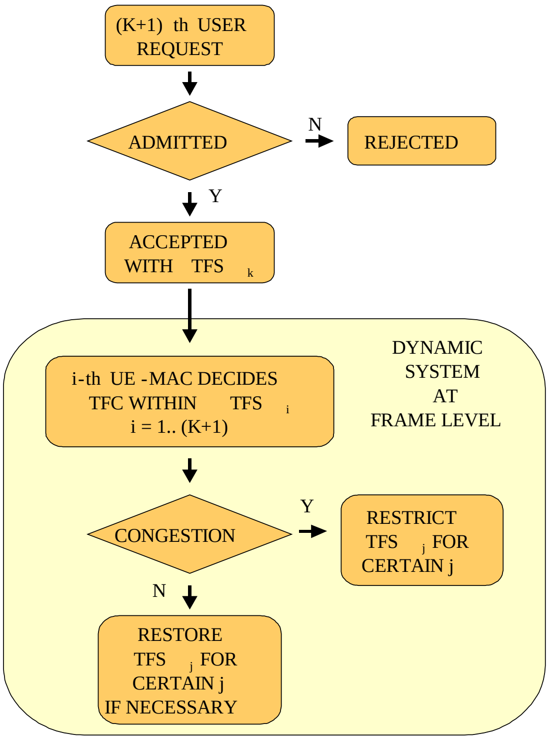
(K+1) th USER REQUEST
ADMITTED
N
REJECTED
Y
ACCEPTED WITH TFS k
DYNAMIC SYSTEM AT FRAME LEVEL
i-th UE -MAC DECIDES TFC WITHIN TFS i
i = 1.. (K+1)
CONGESTION
Y
RESTRICT TFS j FOR CERTAIN j
N
RESTORE TFS j FOR CERTAIN j IF NECESSARY

Figure 1. Uplink RRM approach

Figure 1 summarizes the uplink RRM approach presented in the above subsections. Assuming K users already admitted into the system, the (K+1)th request must pass the admission control phase. If positive, the user is accepted with certain transmission parameters, basically in the form of defining the maximum allowed transmission rate (a TFS is defined). With the allowed range of rates, the algorithm implemented at UE-MAC autonomously chooses the transmission rate each TTI. The RNC has to monitor the current cell load in order to detect possible congestion situations. If congestion is triggered some actions are taken in order to reduce the cell load, usually in the form of reducing the transmission rate capabilities to a certain set of mobiles. When the congestion is passed, the network should restore the original transmission capabilities to the terminals affected during the congestion status.

3.- SYSTEM MODEL

The system simulation model considers a radio access bearer for supporting the interactive service with a maximum bit rate of 64 Kbps in the uplink [6]. TB error rate target is 0.5%. Possible transport formats are detailed in Table 1. 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 simulation model includes a cell with radii 0.5 km, perfect power control is assumed for CDMA interference characterisation and intercell interference is represented by f=0.6. Physical layer performance, including the rate 1/3 turbo code effect, is taken from [7]. The mobility model and propagation models are defined in [8], taking a mobile speed of 50 km/h and a standard deviation for shadowing fading of 10 dB.

Table 1. Transport formats for the interactive RAB.

TrCH type		DCH
TB sizes, bit		336 (320 payload, 16 MAC/RLC header)
TFS	TF0, bits	0×336
	TF1, bits	1×336 (16 Kb/s, SF=64)
	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

4.- RESULTS

In order to gain more insight into the congestion procedure, let consider the following algorithm, where the impact of the different parameters involved will be dealt along this section.

1) Congestion detection algorithm: : if $\eta \geq \eta_{CD}$ in 90% of the frames within ΔT_{CD}

2) Congestion resolution.

a) Prioritisation: In this paper all the users belong to the same service class and have the same requirements, thus having the same priority. Then, the criterion followed to order them depends on the TF they are using at the time when congestion is triggered (the higher the TF, the higher the position in the table).

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.

- Algorithm 1: The user is 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 protocol message "Transport Channel Reconfiguration".
- Algorithm 2: The TFS is limited to TF2, so that users are not allowed to transmit at more than 32 Kbps whereas in normal conditions the maximum rate is 64 Kbps. Similarly, this is carried out through the layer 3 RRC 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 $\eta \le \eta_{CR}$ in 90% of the frames within ΔT_{CR}.

3) Congestion recovery: A "time scheduling" algorithm so that a user by user restoring approach is considered. That is, a specific user is again allowed to transmit at maximum rate (i.e. a "Transport Channel Reconfiguration" message indicating that TFS includes up to TF4 is sent). Once this user has emptied the buffer, another user is allowed to recover the maximum rate and so on.

Comparisons for the two presented load reduction algorithms are summarized in Tables 2 and 3. The performance figures are the admission probability (i.e. the probability that a user request is accepted into the system), the percentage of time while the network is congested and the delay distribution of the transmitted packets during the congestion period. It can be observed that the "softer" load reduction action for Algorithm 2 leads to a more time in congestion and, consequently, a reduction in the admission probability (notice that the first action in congestion is to reject all connection requests). It seems that the firmer actions taken by Algorithm 1 recalls in shorter congestion periods. Additionally, one of the expected impacts of the congestion status is a delay degradation due to the transmission rate capabilities limitation. It can be observed in Table 3 that Algorithm 1 provides a nicer delay distribution compared to Algorithm 2, specially for the 95% percentile.

It is also of interest to study the sensitivity to the detection and resolution thresholds. Thus, Tables 4 to 6 show performance results for two different options: thresholds for an "early" congestion detection and a conservative resolution (represented by η_{CD}=0.8 η_{CR}=0.7) and a representative case for a "late" detection and an optimistic resolution (represented by η_{CD}=0.9, η_{CR}=0.8). The delay distribution appears somehow nicer for the "early" detection, especially for the 50% percentile case. However, the low thresholds leads to more close congestion situations, as shown in the cumulative distribution of the time between congestions and more time in congestion. This means that a higher signaling load would be necessary for the "early" detection case.

Table 2. Results for η_{CD}=0.8, η_{CR}=0.7, ΔT_{CD}=10, ΔT_{CR}=10

	Algorithm 1 (TF0)		Algorithm 2 (TF2)	
Number of www users	Admission probability	Time in congestion (%)	Admission probability	Time in congestion (%)
600	1	≈ 0	1	≈ 0
650	1	≈ 0	1	≈ 0
700	1	0.13	0.99	0.57
750	1	0.33	0.97	2.34

Table 3. Results for η_{CD}=0.8, η_{CR}=0.7, ΔT_{CD}=10, ΔT_{CR}=10, 700 users

Packet delay percentiles during congestion periods	Algorithm 1	Algorithm 2
50%	<0.12 s	<0.16 s
75%	<0.84 s	<1.12 s
95%	<2.94 s	<6.62 s

Table 4. Results for ΔT_{CD}=10, ΔT_{CR}=10

Packet delay percentiles during congestion periods	η_{CD}=0.8 η_{CR}=0.7 ("early")	η_{CD}=0.9 η_{CR}=0.8 ("late")
50%	<0.12 s	<0.2 s
75%	<0.84 s	<0.98 s
95%	<2.94 s	<3.14 s

Table 5. Results for ΔT_{CD}=10, ΔT_{CR}=10

Cumulative probability of the time between congestions	η_{CD}=0.8 η_{CR}=0.7 ("early")	η_{CD}=0.9 η_{CR}=0.8 ("late")
<1 s	20%	3%
< 100 s	60%	12%
<1000 s	99%	48%

Table 6. Results for ΔT_{CD}=10, ΔT_{CR} =10

Number of www users	η_{CD} =0.8 η_{CR} =0.7 Time in congestion (%)	η_{CD} =0.9 η_{CR} =0.8 Time in congestion (%)
600	≈ 0	≈ 0
650	≈ 0	≈ 0
700	0.13	0.02
750	0.33	0.07

Finally, it is also of interest the study the impact of the observation time for triggering congestion actions. Thus, Tables 7 and 8 compare the case of two different congestion resolution windows. According to Table 7, the case ΔT_{CR} =100 which represents a more safe congestion overcome decision leads to longer congestion periods and, consequently, lower admission probabilities because the system is blocked during congestions. In terms of time between congestions, the safe margin only achieves to reduce very close congestion situations while the probability for longer periods is almost the same.

Table 7. Results for η_{CD} =0.8, η_{CR} =0.7,

	ΔT_{CD}=10, ΔT_{CR} =10		ΔT_{CD}=10, ΔT_{CR} =100	
Number of www users	Admission probability	Time in congestion (%)	Admission probability	Time in congestion (%)
600	1	≈ 0	1	0.1
650	1	≈ 0	1	0.32
700	1	0.13	0.99	0.95
750	1	0.33	0.98	2.34

Table 8. Results for η_{CD} =0.8 η_{CR} =0.7

Cumulative probability of the time between congestions	ΔT_{CD}=10, ΔT_{CR} =10	ΔT_{CD}=10, ΔT_{CR} =100
<1 s	20%	4%
< 100 s	60%	58%
<1000 s	99%	99%

5.- CONCLUSIONS

In the framework of Radio Resource Management strategies for W-CDMA systems, this paper has proposed new congestion and admission control algorithms to deal with the peak and average interference in the uplink together with a decentralized UE-MAC mechanism that selects the transport format to keep a certain bit rate. The overall system behavior has been analyzed for different congestion detection and resolution mechanisms. Results reveal that in general firmer actions (i.e., high reductions of the TFS) and early congestion detection algorithms lead to shorter congestion periods with smoother delay degradations.

6.- 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.

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

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

[6] 3G TS 34.108 v.3.2.0, “Common Test Environment for User Equipment. Conformance Testing”

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

[8] 3GPP TR 25.942 v.2.1.3, “RF System Scenarios”