Scheduling with Quality of Service Constraints for real-time and non-real-time traffic in W-CDMA

Angela Hernández Solana, Antonio Valdovinos Bardají*, Fernando Casadevall Palacio***

* Department Electronics Engineering and Communications. University of Zaragoza. (Spain) anhersol@posta.unizar.es
** Department of Signal Theory and Communications. Polytechnic University of Catalonia (UPC); Barcelona (Spain)

Abstract

This paper presents a comparative analysis of the performance of different scheduling strategies to guarantee the Quality of Service requirements of real-time and non-real-time packet data services. A Dedicated Channel is assumed for all transmissions. Over that, centralized demand assignment protocol is implemented in order to guarantee QoS requirements. Effects of power control errors are considered over the performance of the different strategies. Two alternatives are evaluated in transmission over dedicated codes: 1) all users have requested to transmit maintain DPCCH although they don't have permit to transmit and 2) DPCCH is present only if users are transmitting on DPDCH.

1. Introduction

In the last few years, the demand for multimedia and packet data services, based on Internet standards, has dramatically increased in wired networks. This will clearly impact on the design of the third generation radio systems, such as the UMTS promoted by ETSI, where is required to be provided service quality for multimedia communications, mainly Internet access and video/picture. Because many multimedia applications are packet-oriented, it will be essential to optimize third-generation techniques for supporting variable bit rate and packet capabilities with quality of service requirements. Circuit switching should also be supported for the provision of some constant bit rate services or very high quality voice transmission.

A dual mode packet transmission scheme is envisaged within the W-CDMA concept defined in the UMTS terrestrial radio access (UTRA): Common and Dedicated transmission channels. The common channel packet transmission on RACH is typically used for the transmission of short infrequent packets, while dedicated channel with closed power control is used for large packet transmissions. UMTS assigns a fraction of the available bandwidth to a number of shared physical channels of assigned transport formats (Common Packet Channel-CPCH), which are reserved to mobiles stations for a short number of successive frames by means of MAC signaling exchange. Transmission on CPCH has three phases of contention: random access phase and contention resolution phase and transmission. Although power control is allowed in CPCH transmission, on the CPCH is needed to have a restriction on maximum duration. The procedure of CPCH access is described in [1]. For large and frequent data packet, transmission on dedicated channel is considered. Maintenance of channel could be desirable although mobile has no packets to transmit in order to reduce contention access over common channel.

The scope of this study is limited to the dedicated packet transmission where a demand assignment protocol is proposed in order to guarantee service multiplexing with Quality of Service requirements. In that case, the MS sets up a dedicated code using an initial Random Access request, wherein the type of traffic to be transmitted is specified. Then, the network evaluates the request and decides if the necessary resources can be assigned to the MS. Once the Dedicated Channel is assigned to the MS, the channel is still not allowed to start a transmission. It needs to wait until the network specifies the transport format and the time in which it can initiate the transmission. This procedure will introduce some overhead and delay, which can be as smaller as far the length of packets increases. However, in the Dedicated Channel mode the data transfer is more reliable due to the closed loop power control performed and the absence of collisions. In this paper we not consider the multi code transmission. Multirate is performed varying spreading factors.

In the context of W-CDMA, the issue of the Quality of Service (QoS) support can be divided into two subjects: Temporal transparency and semantic transparency. The semantic transparency is obtained by means of the power control while temporal transparency is achieved using transmission scheduling. The time scheduling scheme implemented at the base station is responsible for arranging transmission of packets within their specified rate requirements and delay tolerances while the purpose of power control criterion is to meet the Bit Error Rate (BER) of simultaneously transmitted packets. In any case, the basic approach for packet transmission over dedicated channels is to exploit the property of delay tolerance, characteristic of many data sources, to improve the system efficiency. The main idea is to schedule the transmission of the packets coming from delay tolerant users so as to reduce, at any instant, the interference seen by the other users. As a result, the users can transfer information at higher rates, leading to an overall increase in throughput.

The present study considers channels DPDCH and DPCCH. Two alternatives are evaluated in transmission over dedicated codes: 1) all users have requested to transmit maintains DPCCH although they don't have permit to transmit and 2) DPCCH is present only if users are transmitting on DPDCH. We consider, in both alternatives, the impact of imperfections in the power control mechanism over the system performance, modeling variation of the target E_b/N_o as log-normal random variables.

This work has been supported by the grants TIC99-0941, TIC98-0684 from the Spanish Government and FEDER 2FD97-1070.

2. Minimum Transmitted Power Criterion

Each user has a quality of service requirement that specifies in terms of: maximum transfer delay and delay jitter, guaranteed minimum transmission rate r_i , and maximum bit error rates (BER) or frame error rates (FER) mapped into an equivalent E_b/N_o constraint denoted by γ_i,.

$$\left(\frac{E_b}{N_o}\right)_i = \frac{W}{r_i}\frac{P_i}{\sum_{j\neq i} P_j + \eta_o W} \geq \gamma_i \qquad i = 1,...,N \qquad (1)$$

The base station scheduler is responsible for providing both bounded delay and fair sharing of the available wireless resources while the aim of a defined power control criterion is to meet the Bit Error Rate (BER) requirements of simultaneously transmitted packets.

Given a set of requirements, we adopt as optimization criterion to assign an optimum level of the transmitted powers for all the users in such a way that their sum is minimized, guaranteeing that the E_b/N_o requirements of all of the users are met. This criterion minimizes the interference caused to other cells increasing the system efficiency. It was shown that for a bandwidth W and N transmitter users the power control problem in a cell is feasible if and only if [2]:

$$\sum_{i=1}^{N} \frac{1}{\frac{W}{r_i \gamma_i}+1} < 1 \quad \mathrm{i = 1..N} \qquad (2)$$

If this condition is satisfied for a set of rates and E_b/N_o values, then the power can be obtained using (3):

$$P_i = \frac{\eta_o W}{\left(\frac{W}{r_i \gamma_i}+1\right) C_{res}} \qquad \mathrm{C_{res}} = 1 - \sum_{i=1}^{N} \frac{1}{\frac{W}{r_i \gamma_i}+1} \qquad (3)$$

This inequality is for the case when there are no transmit power limits an only one cell is considered, but it can easily be extended to the transmitted power constrained case and also cell environment.

In any case, as can be seeing, perfect power control is required in order to satisfy inequality. If the inequality is not satisfied, then such a power assignment does not exist and the Eb/No requirements of all users cannot met. Power control loops can be designated to adjust the power of user on an individual basis, based on current conditions for that user. However, power control imperfections have to be considered in order to analyze the real performance.

Now, taking into account the minimum total transmitted power criterion we can consider two transmission modes for delay tolerant users.

- In the first one, all users admitted in the system are allowed to transmit information, with a rate as higher as the allowed in the system to satisfy minimum power constraint requirements.
- For the second case, in a given time instant, only a limited number of users are allowed to transmit while the remaining users can not transmit even though they are in contact with the base through a control channel.

Given that in W-CDMA only a set of spreading factors (W/r_i) can be used, only the second option will be analyze in this paper, because the first option needs infinite granularity in the selection of the spreading factor if the power assignments should be optimized.

On the other hand, W-CDMA defines two types of dedicated physical channels: the dedicated physical data channel (DPDCH), used to carry dedicated data, and dedicated physical control channel (DPCCH), used to transmit control information [pilot bits, transmit-power control (TPC) commands, and optional transport format indicator (TFI)]. The DPCCH is transmitted continuously at a constant symbol rate and spreading factor of 256, with relatively low power and enabling physical maintenance (i.e. closed-loop power control, time synchronization, and up-link channel estimation for coherent demodulation). In the up-link, the DPDCH and DPCCH are transmitted in parallel in phase and quadrature-phase branches, respectively, using different orthogonal codes. Although there is no self-interference among DPDCH and DPCCH, we must consider the effect of the DPCCH channel coming from users. Thus, the condition we must satisfy in order to allow the transmission of M<N users in a given time-slot is:

$$\sum_{i=1}^{M} \frac{1}{\frac{W}{r_{d,i}\gamma_{d,i}(1+m)}+1} + \sum_{j=M+1}^{N} \frac{1}{\frac{W}{r_{c,j}\gamma_{c,j}}+1} < 1 \, ; \quad \begin{cases} \mathrm{M=1..N} \\ m = P_{c,i}/P_{d,i} \end{cases} \qquad (4)$$

wherein r_d, r_c are the rates, γ_d, γ_c are the E_b/N_o constraints for DPDCH and DPCCH respectively, and m is the amount of overhead introduced by DPCCH.

3. Protocol Description.

A centralized demand assignment protocol is implemented in order to guarantee QoS requirements. Every user who has packets waiting for transmission sends a request over Random Access Channel to setup a dedicated code. This initial Random request includes the type of traffic and the amount of data to be transmitted. Once, the dedicated channel (code) is assigned, users wait the notification of the base station that they can transmit in the next frame. At the end of each frame, the base station specifies the set of services allowed to transmit simultaneously, together with the transfer format (e.g. the bit rate) to be used for packet transmission. This procedure is done in conjunction with power control in such a way that the QoS requirements of all scheduled services are met. If user has more packets to transmit the mobile station send a new access request on the dedicated channel or piggybacking information in the last packet transmitted. Useful information of access request is contained at the beginning of frame to allow the response of the base station at the end of the same frame. This access request is performed with maximum spreading factor (W/r=256) in order to decrease interference originated over the rest of users. This is the only amount of load that is not controlled by base station, so, a minimum capacity is reserved for request in order to agree with minimum transmitted power criterion. To increase the data throughput and decreasing delay, data rates can be increased during periods of low activity. Processing delay of a frame is assumed for all packets. The retransmission strategy used is the selective repeat ARQ scheme with negative acknowledgements.

4. Scheduling Disciplines.

In ETSI WCDMA the scheduling is a resource allocation function closely connected to the transport format selection (rate of the dedicated channel, coding used, etc). During communication MAC scheduler selects the appropriate transport format within an assigned transport format set for each active transport channel depending on source rate and radio resource limitations. The selection can be done on a 10ms frame basis or slower. Depending on the selected transport format one or more transport block can be transmitted. The main objective of the scheduler is to integrate traffic sources with different transmission rates, priorities, delays and packet loss requirements optimizing the uplink channel utilization. We propose and evaluate several scheduling strategies based both in static or dynamic priorities in order to ensure QoS requirements in terms of rate and minimum delay. Between service classes static priorities are used. We contemplate several possibilities:

- All resources are available for all classes of traffic with or without establishment of different levels of priorities between classes.
- A minimum capacity is guaranteed for non-delay constrained traffic while the remaining is assigned with preemptive priority to delay constraint traffic.

For all packets belonging to the same class, dynamic priorities based on lifetime of packets (Time Stamp Strategy) are applied and compared with Round Robin. Priorities based in lifetime, are calculated as:

$$vt_k \leftarrow max(d_i - (t - t_a), vt_{k-1} + T) \qquad (5)$$

The lifetime (in frames) of the packet placed on the first places of queue of terminal, when it sends a request for transmit, is calculated as d_i-(t-t_a),, where d_i is the delay tolerance of packet (normalized to the frame duration) and t and t_a are respectively the current frame number and the frame number when packet was generated. The lifetime of the other packets is calculated according to equation (5), where T is the estimated packet inter-arrival time. The base station has a request table containing terminal requirements and the lifetimes of next packets to be transmitted. The lifetimes are updated and decreased at each frame. The base station has not to be informed about the arrival of each new packet because it can estimate the time of the next packet applying the same scheme. Only in case of a faulty estimation, the wireless terminal has to transmit an explicit capacity request in order to resynchronize the estimation algorithm. The packets to be transmitted are scheduled in increasing order of lifetime. If vt>d, the next packet to be reserved has not yet been generated. Therefore the reservation will be inserted in service queue as soon as vt=d. Packets with delay constraints are retransmitted until they are correctly received, or their deadlines are violated. As a consequence of the dynamic frame assignment, an error packet can be recovered immediately through a retransmission attempt. In round robin strategy a compensation algorithm is implemented when users, which are allowed to transmit has no enough resources and remaining capacity is assigned to the next user in the round. In any case, when all users have received its corresponding service, remaining capacity is shared increasing the rate of users with packet waiting to be transmitted.

We will present the results of simulation experiments that illustrate the performance of the scheduling algorithms in terms of average and maximum delay, throughput and packet loss rate. When service degradations occur, and in order to evaluate how fairly they are distributed among the users that belong to the same traffic class, we assess not only the mean values but also the distributions of QoS parameters.

5. Traffic Source Models.

The paper considers two kind of traffic sources:

- Real Time Services (Class I and II). Data services with low delay constraints (50ms) and long delay constraints (300ms) respectively and a Block Error Rate (BLER)<10^{-2}.
- Non-real Time Services (Class III). Data services with non-delay constraints and a Block Error Rate (BLER)<10^{-2}.

Convolutional coding rate ½ together with a retransmission scheme (ARQ) are used to achieve BLER=10^{-2}.

Although circuit-switched mode transmission has been proposed by ETSI for Class I and II, packet–switched and channel activity<100% have been considered in this work, in order to support multimedia real time services.

Therefore, in our simulations, real-time sessions are based on Packet Calls with a number of packets exponentially distributed with mean 35 packets, while a service of 36kps (transport block of 360bits) is assumed although real transmission rate is 128kbps, so a spreading factor of 32 is used. Average inter-packet arrival time is 10ms while packet call inter arrival time is exponentially distributed. From the viewpoint of non-real time sessions, the traffic source is based in model presented in [3]. In particular 8kbps data rate is assumed. Sessions consists of a sequence of packet calls during which several packets may be generated. Normal Pareto distribution with mean 480 bytes is considered for the data packet size while an inter arrival time between packets of 500ms is assumed, being 25 the average number of packet within a packet call. Average time between the last packet of a packet call and the next packet call is 4s. Packets are segmented in PDU contained in transport blocks of 360 bits.

Power ratio of –3.59 dB between DPCCH and PDDCH and static propagation condition with additive white Gaussian noise are considered as is propose to be used in [4]. These conditions give a required Eb/No=1.75dB (considering overhead of DPCCH) to achieve a BLER=10^{-2} assuming an interleaving of 10ms.

The return channel (downlink DPDCH-PDCCH) is assumed to be error free.

6. Simulation Results.

In this subsection simulations results are presented, all of them considering both DPDCH and DPCCH channels.

Figures 1 to 5 show a comparative performance of Round Robin and Time Stamp scheduling considering only one type of traffic, particularly mobiles of class II, and the two alternatives mentioned early about transmission over dedicated codes:

1) All users have requested to transmit maintains DPCCH although they don't have permit to transmit.
2) DPCCH is present only if users are transmitting on DPDCH.

The impact of imperfections in the power control mechanism over the system performance is also evaluated, modeling variation of the target E_b/N_o as lognormal random variables with different standard deviations.

In figures 1 to 4 the packet call inter arrival is 1s while in figure 5 packet inter arrival is 500ms.

Figure 1 shows packet dropping probability versus number of simultaneous users per cell, with DPCCH always present (Closed Control), and results with power control imperfections of 0.5,1, 1.5 and 2.5dB. Figure 2 shows this probability with DPCCH only present in transmission. In this case, for all the curves, a variation of 2.5dB is considered for the first transport channel (10ms) of a transmission burst while variations of 0.5, 1, 1.5 and 2.5dB are considered for the rest of the transport where closed control is acting (curves are denoted by 2.5dB+xdB). As quality criterion we consider a maximum tolerated dropping probability equal to 1%. These two figures will permit us to compare divers aspects. First, we evaluate independently the performance of Round Robin and Time Stamp. Comparing, closed control with a given power control imperfection of x dB with the corresponding curve with 2.5dB+x dB, we see that the number of mobiles that can be supported is higher in all cases when only DPDCH is present while users are transmiting on DPDCH.

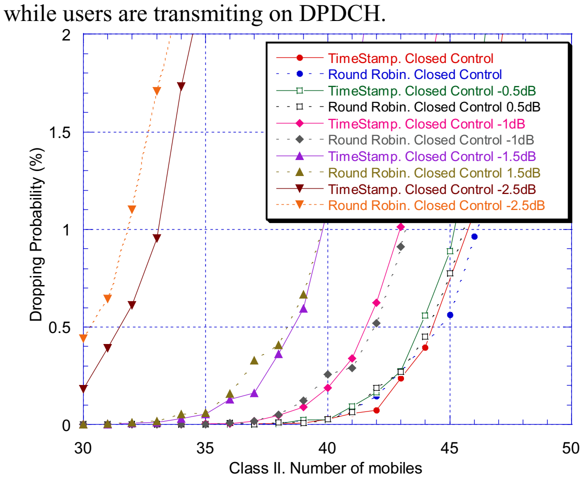

Figure 1

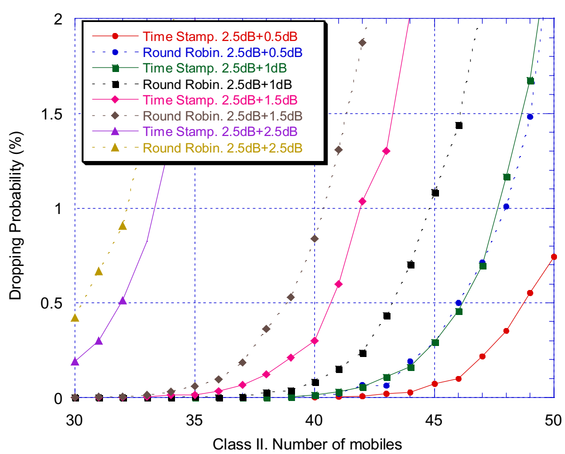

Figure 2

The more realistic comparison has us to note that even perfect closed control and closed control with little variations of 0.5dB offers worse results than 2.5dB +0.5dB. Then, comparing Time Stamp with Round Robin in closed control, we can see that Round Robin offers the best results for perfect closed control and little variations. This is because users spend few time waiting to transmit after they request and in consequence waste of capacity in DPCCH is smaller. But comparative performance of Time Stamp improves as errors in closed control increase. For 1.5dB results are similar to Round Robin while for 2.5dB Time Stamp performs better compensating the effect mentioned before. So we can conclude that Time Stamp offers a more robust performance. Seeing results for 2.5dB+x dB we can see that Time Stamp offers the best results.

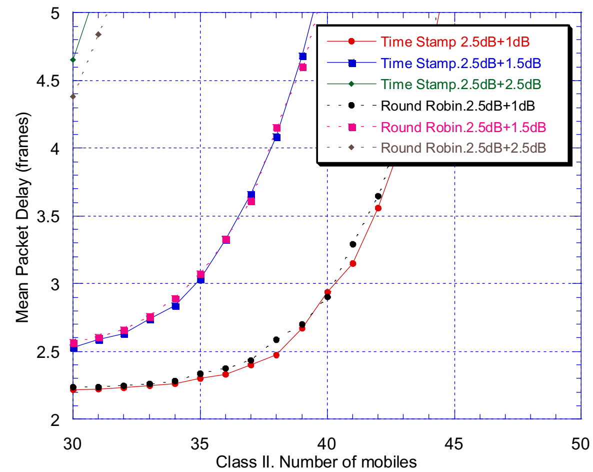

Figure 3

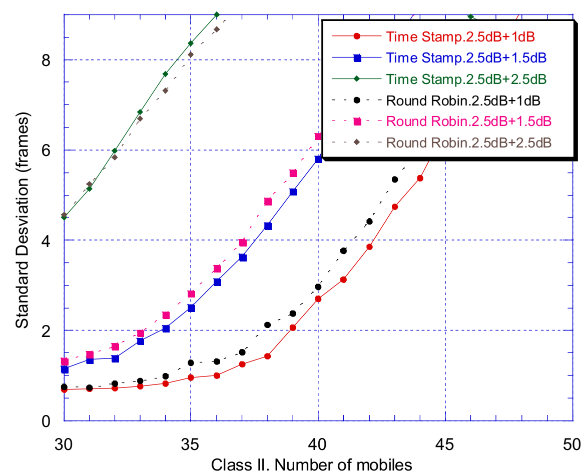

Figure 4

Figure 3 shows mean packet delay for 2.5dB+xdB cases. As can be seen, differences in packet delay are not as significant as packet dropping probability. However, we have to be present that although mean packet delay is similar only a number of users that guarantee dropping probability less than 1% could be supported. Figure 4 shows standard deviation of packet delay. As can be seen, for the margin of interest standard deviation is fewer for Time Stamp strategy. Note that for mean packet delay and standard deviation, packets dropped are no considered in statistics.

Finally, figure 5 shows a comparative between Time Stamp and Round Robin for perfect closed control and 2.5dB+0.5 dB when inter arrival packet is 500ms. Results are quite similar to the obtained when an inter arrival packet of 1s is assumed.

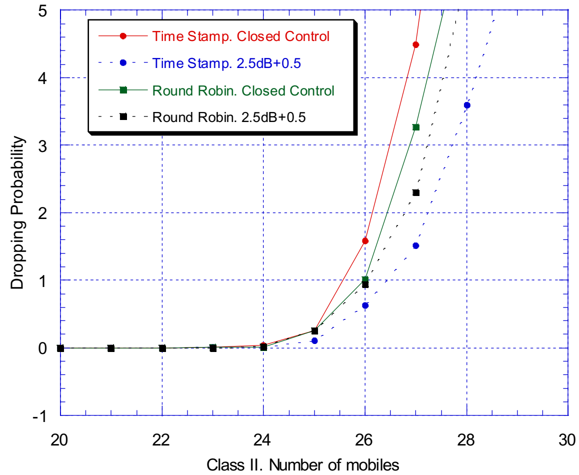

Figure 5

Figures 6 to 9 show the behavior of the system considering the three different services simultaneously and perfect closed control. 15 class I and 15 class II mobiles are considered with packet inter-arrival 1seg. Among packet belonging to the same class dynamic priorities based on lifetime of packets are applied and compared with Round Robin. Also we compare 3 alternatives in the management of different services: all resources available for all classes and two priority levels with and without reserve of a minimum of capacity of 10% for non-real time services. In particular, figures 6 and 8 shows dropping probability for the two real time services while figures 7 and 9 shows mean packet delay for not real time services, for Round Robin and Time Stamp respectively. Results for the two strategies are significantly different. For Round Robin strategy, considering only one priority level offers the best results since the number of users waiting to transmit decrease significantly. However, in Time Stamp real time services are delayed in order to improve real time performance so users waiting to transmit decreasing global capacity in order to maintain their channel DPCCH. We can appreciate those in that case two priority levels offers best results. It has to consider that dropping probability for class I determine the maximum load allowed in the system. In any case Time Stamp performance is worse than Round Robin.

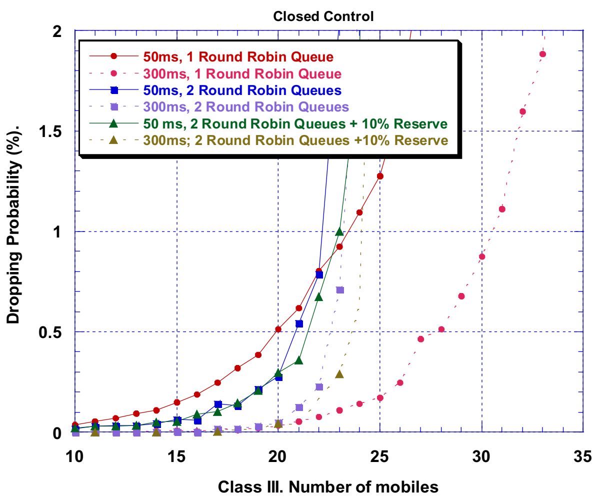

Figure 6

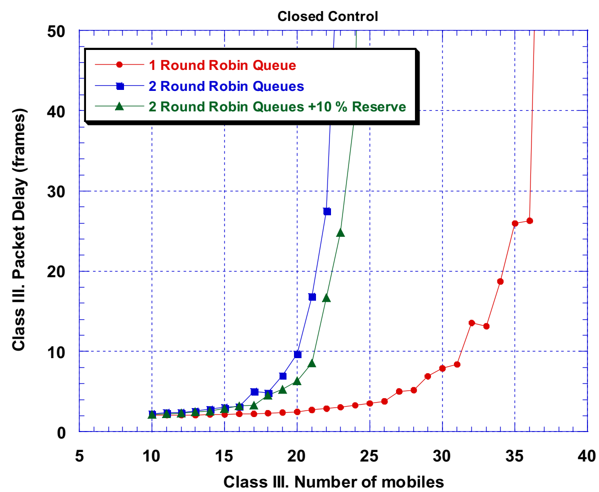

Figure 7

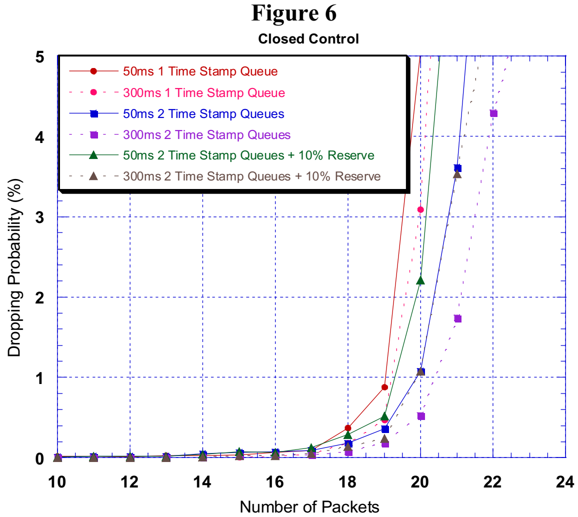

Figure 8

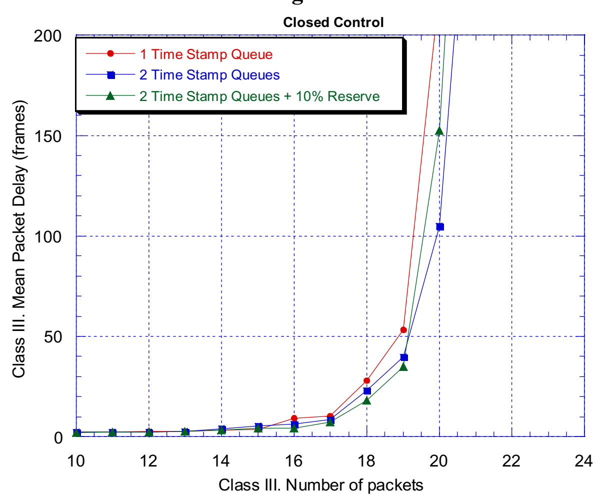

Figure 9

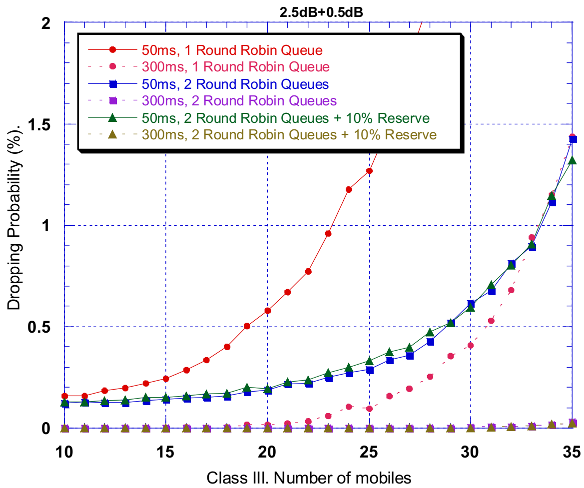

Figure 10

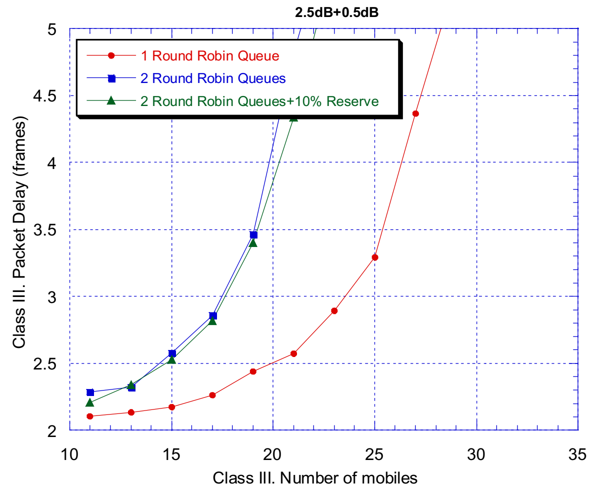

Figure 11

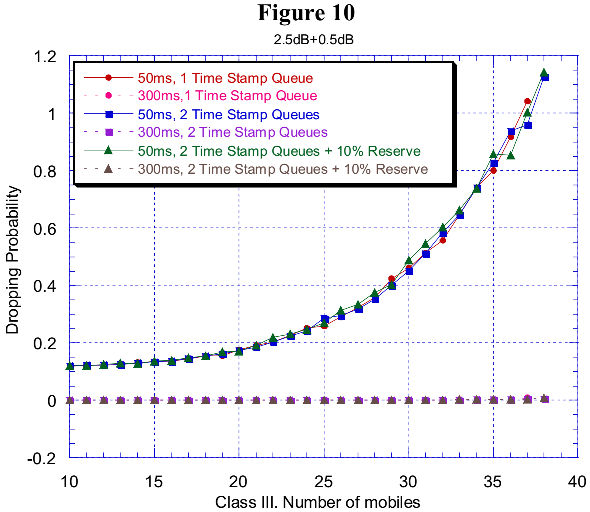

Figure 12

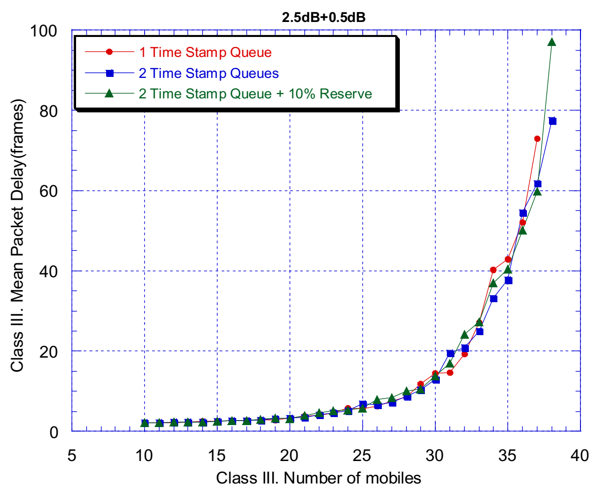

Figure 13

Figures 10 to 13 shows the same parameters when DPCCH is present only if users are transmitting on DPDCH. Results improve and differ significantly from the case of perfect closed control. In spite of the fact that number of users supported increase, it's important to note that dropping probability for Class I service increase when load of system is still low. In Round Robin case (figures 10 and 11) better results are achieved when two priority levels are considered. 33 users of class III can be supported in front of 23 when closed control and only one priority is considered. While, Time Stamp offers a performance must better than Round Robin in any case (37 users of class III). So, Time Stamp results to be more robust in front of worse conditions as we have concluded before. No obvious differences are contemplated between considering all the classes with the same priority and considering two priority classes with and without reserve of resources for non real time services. Note, that delay obtained for non real time services is high as we can expect due that better performance of real time services is possible delaying non real time services.

7. Conclusions

This paper presents and evaluates several scheduling strategies for packet transmission in the FDD mode of UMTS. These strategies are based on the establishment of different levels of priorities between classes of traffic combined with assignment of dynamic priorities associated with individual packets. Performance of scheduling strategies shows dependent of the efficiency in power control. Additionally, it can be conclude also that presence of DPCCH when mobile is not transmitting on DPDCH reduce the efficiency of scheduling strategies.

8. References

[1] 3GPP TS 25.321 " MAC Protocol Specification", 3G TS 25.321 v3.6, December 2000.

[2] Ashwin Sampath, Sarath Kumar and Jack M. Holtzman, "Power Control and Resource Management for a Multimedia CDMA Wireless System", PIMRC'95.

[3] "Universal Mobile Telecommunications System (UMTS). Selection procedures for the choice of radio transmission technologies of the UMTS", UMTS Technical Report 30.03, version 3.2.0. April 1998.

[4] 3GPP TS 25.321 " UTRA(BS) FDD; Radio transmission and Reception", 3G TS 25.104 v3.5.0, December 2000.