Scheduling with QoS Support for MultiRate Wireless Systems with Variable Channel Conditions – Extended Version

Alexandros Gougousis and Michael Paterakis
Information and Computer Networks Laboratory
Dep. of Electronic and Computer Engineering
Technical University of Crete
73100 Chania , Crete
{xayate,pateraki}@telecom.tuc.gr

***Abstract*— Scheduling through wireless shared channels is a very challenging problem because of each user's link-state variations through time. The challenge lies in the support of the quality of service (QoS) of each user under the uncertainty of channel behavior. This problem becomes even more difficult for some of the 2.5G and 3.5G network architectures (e.g EDGE,HSDPA, HDR) where the supported transmission rate can take multiple values and the channel can not be modeled by the traditional ON-OFF model. In this paper we propose Expo-pred-wei, a new scheduling algorithm, which take advantage of the channel variations in order to maximize the total cell throughput and at the same time to preserve, in an efficient way, a minimum QoS for each user. In contrast to other proposed algorithms, our algorithm exhibits very good performance in a wide set of traffic scenarios including Real-Time (RT) or Non-Real-Time (NRT) traffic as well as mixed RT and NRT traffic.**

***Index Terms* — Scheduling, multirate wireless channel, QoS support, HSDPA**

I. Introduction

One of the most important issues emerging with the development of packet switching wireless cellular networks is scheduling through downlink shared channels with varying condition. Until now, wireless networks were, in their majority, supporting only a unique transmission rate and they have been being modeled as an ON-OFF channel [1],[2],[3], expressing if a transmission could be successful under the current channel condition. With the introduction of new techniques like Adaptive Modulation and Coding (AMC) in cellular networks (e.g EDGE [4], HDR [5], HSDPA [6], 802.16 [7]), the states in which a radio-link can be (i.e the transmission rate that a link can support with certain BER) are multiplied. The user mobility in such systems leads to variations of each user's link state and as a consequence to variations of the maximum transmission rate the user can support every moment. These variations happen in a small (msecs) or bigger (secs or tenths of secs) timescale. The former variations, which are the most interesting because they highly affect the scheduler's performance, are caused by multipath propagation which is a very intense phenomenon in urban environments [8]. Efficient radio resource management in such systems should include a scheduling algorithm that will successfully integrate combinations of different traffic types and at the same time efficiently exploit the link variations.

Existing work includes algorithms that do not into consideration in real-time, the QoS requirements or the current QoS satisfaction of each user (best-effort algorithms [9],[10],[11],[12],[13],[14],[15]) as well as algorithms that embody mechanisms for real-time QoS support (QoS-based algorithms, [16][17][18][19][20]). The former are unable to guarantee a required minimum QoS for each user, while the later exhibit good performance only under certain traffic scenarios

Most of the work on the subject has mainly considered limited traffic scenarios which usually include only RT or NRT traffic and a limited number of scheduling algorithms for comparison. In this work we study the behavior of scheduling algorithms that have been proposed for systems with varying transmission rate for each user, for a wide range of traffic scenarios (RT, NRT and combined RT and NRT) traffic. The simulation environment is a UMTS cell employing HSDPA. We also design and propose a new QoS-based scheduling algorithm (Expo-pred-wei) which exhibits very good performance under all traffic scenarios examined.

The rest of this paper is organized as follows: Section II gives a description of the system model and section III presents the expressions that would be used for the description of the QoS requirements. Section IV includes an overview of the most important existing scheduling algorithms. In Section V we describe the operation of the proposed scheduling scheme. Section VI includes a description of the traffic models used in our simulations. Section VII contains our simulation results and their discussion, while section VIII summarizes the main conclusions of the paper.

II. System Model

The system we will use for the performance comparison of the scheduling algorithms is the HSDPA architecture that has been proposed for UMTS. The time is divided into 2msec timeslots (or TTI, which stands for Transmission Time Interval) and only one user is scheduled for transmission in each timeslot. Besides time multiplexing, transmission to more than one user (e.g 2 to 4 users) in each TTI is possible in HSDPA by using different parts of the code set for each

* Research supported in part through a U.S.A – Greece bilateral collaborative research grant by the Greek Secretariat for Research and Technology.

scheduled user (code multiplexing). In our simulations we consider that multiplexing is taking place only in the time domain (all of the system resources are granted to only one user every TTI) as this is the main scenario under consideration by the most of the researchers in the field.

The HSDPA architecture uses a mechanism that estimates the condition of each user's link. The link variations are simulated through Jake's model for Rayleigh fading channels [21]. User speed is assumed equal to 3 Km/h for every user. A random error that follows normal distribution is introduced in each link-state estimation to emulate the estimation error of the real system. Based on this estimation, at the beginning of each TTI the base station evaluates the best combination of modulation and coding schemes that each user can support for a certain BLER (i.e the maximum transmission rate that each user can support) in this TTI. Taking this information and the buffer state into account, the scheduling algorithm decides which user will transmit in this TTI. Finally, an additional feature of the HSDPA architecture that is being simulated is Hybrid ARQ. In HARQ the information from a previous unsuccessful transmission is combined with the information of a new retransmission in order to increase the probability of correct packet decoding at the receiver. This technique enhances the communication through a bad link by reducing the number of retransmissions required for the correct reception of a packet.

III. DESCRIPTION OF QOS REQUIREMENTS

Here we introduce a description for the QoS requirements of each user. We assume that each flow/service corresponds to only one link and one user. In most relative works QoS requirements are expressed for NRT service as: $R_i \geq r_i$, while for RT service as: $Prob\{D_i > T_i\} \leq \delta_i$, where R_i is the (estimated) transmission rate supported by user i for the current TTI and D_i is the current delay suffered by the Head-Of-Line (HOL) packet of user's i queue. That means that each NRT flow requires a minimum throughput, while each RT flow requires that its packets are transmitted with delay lower than a predefined threshold with a certain probability.

IV. PREVIOUS WORK

Our proposal belongs to the QoS-based algorithms, algorithms that embody mechanisms for better and efficient support of the desired QoS [16][17][18][19][20] by using more complicated scheduling criteria that include more than one scheduling targets (in contrast to Best-Effort algorithms that use only one scheduling target). The main QoS-based algorithms proposed so far are:

M-LWDF (Modified Largest Weighted Delay First) : This algorithm [16][17] is one of the first QoS-based algorithm proposed and schedules user j selected by the following criterion :

$$j = \arg\max_i \left\{ \frac{R_i(t)}{\overline{R_i}} \cdot a_i \cdot D_i \right\}$$

A suggested value for a_i is : $a_i = \frac{-\log(\delta_i)}{T_i}$, (the meaning of δ_i and T_i has been given in section.III) hence the scheduling criterion becomes:

$$j = \arg\max_i \left\{ \frac{R_i(t)}{\overline{R_i}} \cdot \left(-\log(\delta_i) \cdot \frac{D_i}{T_i} \right) \right\}$$

This choice for a_i is analytically derived [22] but also has an intuitive reasoning by that it gives more priority to users with stricter QoS requirements, which are the most sensitive users in periods of bad channel condition.

EXPONENTIAL-RULE : This algorithm [18] attempts to equalize the weighted delays of all users. The scheduling criterion for selecting user j is :

$$j = \arg\max_i \left\{ a_i \cdot \frac{R_i(t)}{\overline{R_i}} \cdot \exp\left(\frac{a_i \cdot D_i - \overline{a \cdot D}}{1 + \sqrt{\overline{a \cdot D}}} \right) \right\}$$

The authors of [18] note that the the mean value in the numerator of the exponent is introduced only for better understanding of the criterion's functionality and the removal of this term doesn't change the functionality of the algorithm. Hence, in some papers Exponential-Rule is mentioned or used in the following form (which we would refer to as Expo-drop algorithm in the rest of this paper):

$$j = \arg\max_i \left\{ a_i \cdot \frac{R_i(t)}{\overline{R_i}} \cdot \exp\left(\frac{a_i \cdot D_i}{1 + \sqrt{\overline{a \cdot D}}} \right) \right\}$$

, where the parameter a_i is chosen as in M-LWDF.

CD-EDD (Channel-Dependent Earliest Due Date) : This algorithm [19], is a modification of M-LWDF and uses the criterion:

$$j = \arg\max_i \left\{ a_i \cdot \frac{R_i(t)}{\overline{R_i}} \cdot \left(-\log(\delta_i) \cdot \frac{D_i / T_i}{T_i - D_i} \right) \right\}$$

, where the parameter a_i is chosen as in M-LWDF

Note.1: For the algorithms mentioned above and in the rest of the paper we assume that R does not denote the current transmission rate that can be supported but instead the signal reception power (which of course, is directly related to the transmission rate that can be supported).

Note.2 : The usage of HOL-packet delay as an indicator of the satisfaction for users that require a minimum throughput is made by using the technique of Token Queues as proposed and described in [17],[20].

V. PROPOSED SCHEDULING SCHEME

Based on our simulations, the QoS-based algorithms presented in section IV exhibit quite good performance when they are applied to systems with only RT or NRT (depends on the algorithm) traffic and under specific traffic, but not under a wide range of traffic scenarios (RT, NRT, RT and NRT). In this section we propose a new scheduling algorithm, named

Expo-pred-wei, that eliminates this problem. The development of Expo-pred-wei is a result of working on some other intermediate algorithms that have been developed during a deeper study of the underlying scheduling process. These intermediate algorithms reveal the motivation of Expo-pred-wei and will be shortly presented in the following.

A. Expo-Linear

Expo-Linear, which is the basic form of Expo-pred-wei, is a modification of Exponential-Rule, at each TTI it schedules user j selected by the criterion:

$$j = \arg\max_i \left\{ a_i \cdot \frac{R_i(t)}{\overline{R_i}} \cdot \exp(a_i \cdot D_i) \right\}$$

, where the parameter a_i is chosen as in M-LWDF

The motivation for the development of Expo-Linear was the problematic behavior of the algorithms presented in section IV, under specific cases of scheduling conditions. Examples of such cases are presented in Table I. High cruciality in Table I means that the delay of the HOL packet is approaching the packet expiration time (i.e the maximum tolerable delay). The removal of the denominator, from the criterion used by Expo-drop, in Expo-Linear, has two advantages. First, it alows the algorithm to increase the total throughput when no user's QoS is in danger. Second, it allows a more direct influence by the service whose HOL packet delay is approaching its deadline since this influence is not anymore affected by the condition of the HOL packets of the other users.

B. Expo-weight

Expo-weight schedules user j selected by the following criterion:

$$j = \arg\max_i \begin{cases} -\frac{\log(\delta_i)}{T_i} \cdot \frac{R_i(t)}{\overline{R_i}} \cdot \exp\left\{-\log(\delta_i) \cdot \frac{D_i}{T_i}\right\} & , \text{ for Non Real Time flows} \\ -\frac{\log(\delta_i)}{T_i} \cdot \frac{R_i(t)}{\overline{R_i}} \cdot \exp\left\{-\log(\delta_i) \cdot \frac{D_i}{T_i}\right\} \cdot RT_{on} \cdot \frac{Drops_i}{\sum_{j \in ON} Drops_j} & , \text{ for Real Time flows} \end{cases}$$

, where RT_{ON} is the number of RT users that have packets for transmission.

In Expo-weight we use a modification of the balancing technique used by CD_EDD in [19]. According to this modification, the mean value of the number of rejected packets (Drops) is calculated and used only for the RT flows which have packets for transmission. The motivation for this choice is explained through the following example. Assume that we have 5 RT flows named A,B,C,D and E with the following drops: 2 packets for each of A,B and C, and 12 packets for each of D and E. The mean value of packet drops for all the RT flows is 6. Let us also assume that only A,B and C flows have packets for transmission. In this case if we use the balancing technique of [19], the priority of flows A,B and C will be reduced without this helping to the balancing of the QoS dissatisfaction between the RT flows. On the contrary, this reduction will bring A,B and C to a disadvantageous position relative to the NRT flows without this being our goal. As a result this technique can lead the scheduler to favor the NRT flows, which, by their nature, are more tolerable to QoS degradation. This behavior is undesirable. In the case of Expo-weight this doens not happen because the mean value will be calculated only for the RT flows A,B and C and it will be equal to 2.

C. Expo-predict

Expo-predict schedules user j selected by the criterion:

$$j = \arg\max_i \left\{ -\frac{\log(\delta_i)}{T_i} \cdot coef\_ER \cdot \frac{R_i(t)}{\overline{R_i}} \cdot \exp\left\{ -\log(\delta_i) \cdot \frac{D_i}{T_i} \right\} \right\}$$

, where coef_ER is given by the following rule : ***IF*** *(flow j is RT flow)* ***AND*** $(T_j - \delta_j) < 4$ ***AND*** *(* R_j *is bigger than the number of HOL packet's bits remaining for transmission)* ***THEN*** *coef_ER = 10* ***ELSE*** *coef_ER =1*

The difference between Expo-predict and Expo-Linear lies in the introduction of a coefficient which takes a high value when there is a HOL packet that expires in a few TTIs and this packet can be transmitted in one TTI with the current supported transmission rate. The motivation for this introduction was the observation that in some cases a notable number of packets that could be transmitted (or finish their transmission) in just one TTI where dropped. The high value that «coef» takes in such situations (otherwise its value is "1") should be high enough to overwhelm the variations of the criterion's terms related to the condition of the channel and the variations due to the balancing functionality

D. Expo-pred-wei

Expo-pred-wei comes from Expo-Linear after incorporating the techniques used by Expo-predict and Expo-weighted and is expressed by the following criterion:

$$j = \arg\max_i \begin{cases} -\frac{\log(\delta_i)}{T_i} \cdot \frac{R_i(t)}{\overline{R_i}} \cdot \exp\left\{-\log(\delta_i) \cdot \frac{D_i}{T_i}\right\} & , \text{ for Non Real Time flows} \\ -\frac{\log(\delta_i)}{T_i} \cdot \frac{R_i(t)}{\overline{R_i}} \cdot coef \cdot \exp\left\{-\log(\delta_i) \cdot \frac{D_i}{T_i}\right\} & , \text{ for Real Time flows} \end{cases}$$

where:

$$coef = \begin{cases} coef\_ER & , \text{ if (Real-Time flow) AND } (Ti - Di < 4) \\ RT_{on} \cdot \frac{Drops_i}{\sum_{j \in ON} Drops_j} & , \text{ if (Real-Time flow) AND} (Ti - Di \geq 4) \text{ AND (therearedrops)} \\ 1 & , \text{ inothercase} \end{cases}$$

VI. NETWORK TRAFFIC MODELS

For our simulations we have chosen three different traffic sources which are presented in the following :

FTP : These users are considered "greedy" as far as their need in throughput is concerned, because they have always data for transmission. It usually represents users that download a big file from the network (usually through the FTP protocol).

WWW : Models web browsing users with bursty arrivals (burtsy users). The meantime between the end of a burst arrival and the beginning of a new one follows a geometric distribution, while the size of a burst follows a truncated exponential distribution with maximum and minimum burst size 100 Kbytes and 2 Kbytes, respectively. The mean value of the distribution is 25 or 30 Kbyte, depending on the case. To make the scenario more realistic, the arrival of a packet is not always completed during one TTI, instead it is considered that the packets are arriving at the base station through a 2 Mbps connection. For example, based in this assumption, the arrival of a 10 Kbyte packet will be completed in 20 TTI.

VIDEO : This source models video users as representatives of the voice/streaming service. The source produces a packet of 5120 bit every 40 msecs (data are produced with a rate of 128 Kbps) but a small random delay jitter ($\leq \pm 6$ msec) is introduced in each arrival relative to the predefined arrival time. With this jitter, the source is better approaching the behavior of a real streaming source since the network always introduces some delay variability to the packet transport.

VII. PERFORMANCE EVALUATION

In this section we evaluate the performance of the algorithms presented in sections IV and V, as well of some modified versions of these algorithms, under RT, NRT or combined RT and NRT traffic scenarios. The modifications are relative to the idea of balancing which is introduced in [19] for such algorithms. The algorithms that are simulated and compared are : 1) Expo-drop, 2) MLWDF, 3) CD_EDD, 4) CD_EDD_balanced : CD_EDD with the balancing technique proposed in [19], 5) CD_EDD_weight : CD_EDD with our balancing technique, 6) Expo-Linear, 7) Expo-predict, 8) Expo-weight, 9) Expo-balance : Expo-Linear with the balancing technique proposed in [19], 10) Expo-pred-wei, 11) Expo-pred-balanced : Expo-predict with the balancing technique proposed in [19].

A.Simulation of NRT traffic scenarios :

SCENARIO #1 : We are evaluating the performance of M-LWDF, CD_EDD, Expo-Linear and Expo-drop in a scenario of 8 FTP users. The mean values of Es/No for each user are: User1 : 3 dB, User2 : 7 dB, User3 : 11 dB, User 4: 15 dB, User5 : 19 dB, User6 : 23 dB, User 7 : 27 dB, User 8 : 31 dB. Required QoS is 64 Kbps throughput, the same for all users. We also have: $\boldsymbol{T_i = 1\,\text{sec}}$ *and* $\boldsymbol{\delta_i = 10^{-2}}$.

Simulation results of scenario #1 are presented in Figures 1-3.

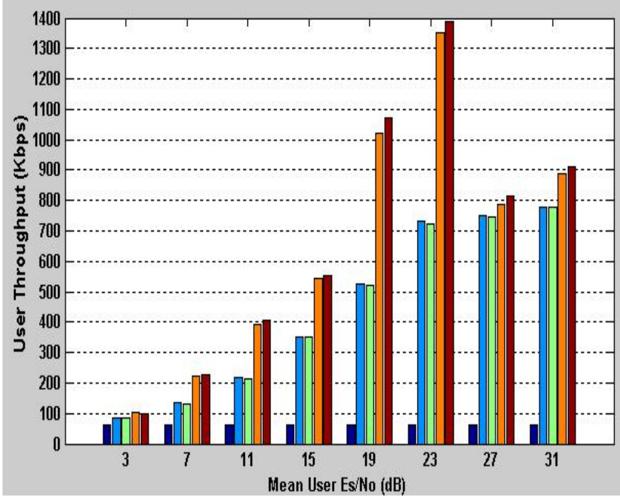

Figure 1 : Achieved throughput by each user for each of the 4 algorithms we examine. Mean Es/No for each user appears on the horizontal axis. The order of algorithms is the same as the one refered in the scenario.

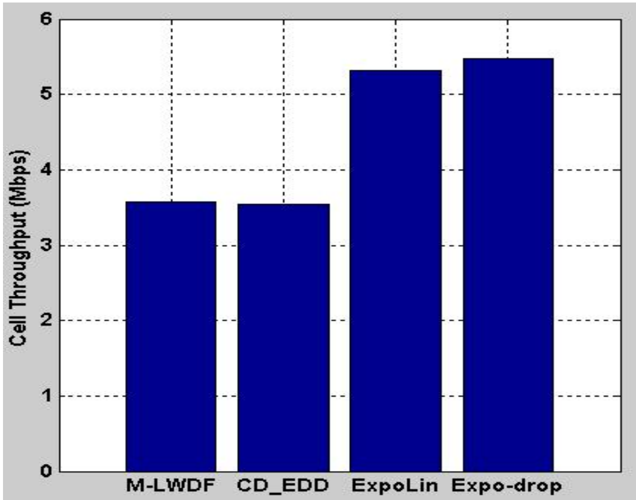

Figure 2 : Total cell throughput achieved by each algorithm.

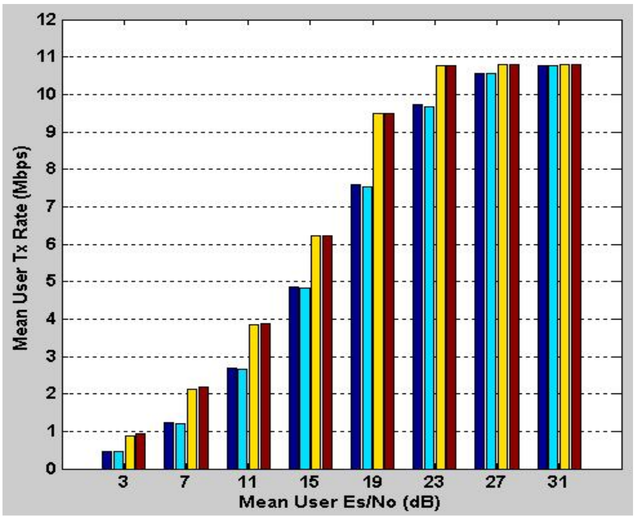

Figure 3 : Mean transmission rate used by each user for each scheduling algorithm.

As we can see, every user achieves throughput higher than the minimum required (Figure1). Expo-Linear achieves as high cell throughput as Expo-drop and their throughput performance is about 50% higher than that of MLWDF and CD_EDD (Figure2). Higher throughputs achieved by Expo-Linear and Expo-drop is due to a better selection of the moments that each user is going to transmit (Figure 3), especially for users that are not permantly or almost permantly in a good link-state (like the two last users in Figure 3).

SCENARIO #2 : We are evaluating the performance of M-LWDF, CD_EDD, Expo-Rule and Expo-drop in a scenario of 9 WWW and 6 FTP users. The mean values of Es/No for each WWW user are: User1 : 3 dB, User2 : 7 dB, User3 : 11 dB, User 4: 15 dB, User5 : 19 dB, User6 : 23 dB, User 7 : 27 dB, User 8 : 31 dB, User 9 : 35 dB, while for the FTP users are: User1 : 5 dB, User2 : 10 dB, User3 : 15 dB, User 4: 20 dB, User5 : 25 dB, User6 : 30 dB Required QoS is 64 Kbps for each FTP user. We also have: $T_i = 1\,\mathrm{sec}$ *andι* $\delta_i = 10^{-2}$ *for FTP users and* $T_i = 4\,\mathrm{sec}$ *και* $\delta_i = 10^{-2}$ *for WWW users.*

Simulation results of scenario #2 are presented in Figures 4-6.

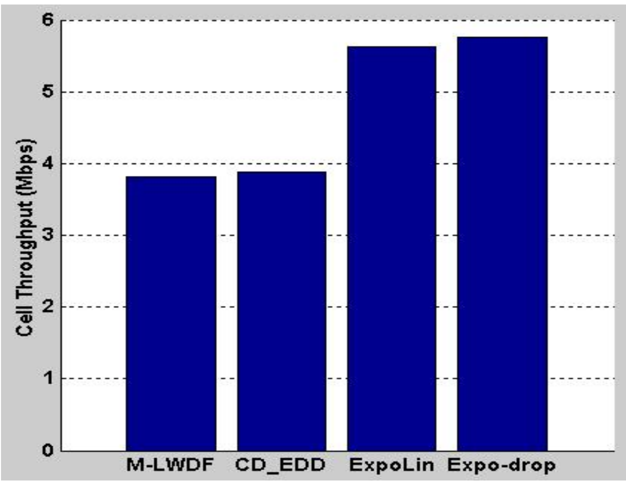

Figure 4 : Total cell throughput for each of the 4 algorithms.

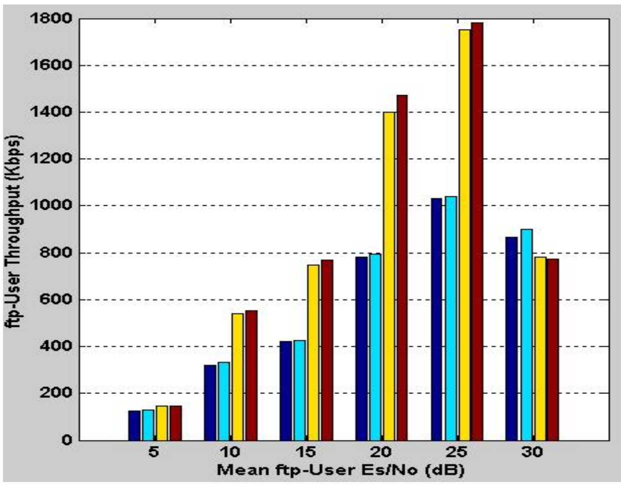

Figure 5 : Mean throughput achieved by each ftp user. Mean Es/No for each ftp user appears on the horizontal axis. The order of algorithms is the same as the one refered in the scenario.

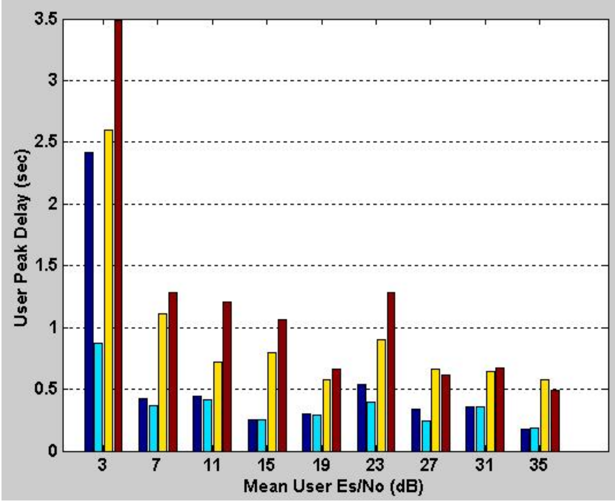

Figure 6 : Maximum packet transmission delay for each WWW user achieved by each scheduling algorithm.

Expo-Linear and Expo-drop achieve about 50% higher cell throughput than that of MLWDF and CD_EDD (Figure 4). This is mainly because of the higher throughputs achieved by FTP users (Figure 5). Low cell throughput of MLWDF and CD_EDD is the price for achieving pretty low delays for the WWW users (Figure 6). Delays achieved by Expo-drop and Expo-Linear are higher but still acceptable. While Expo-Linear's cell throughput is only 2.3% lower that Expo-drop's, Expo-Linear decreases the maximum value of mean and maximum delay among www users by 20% and 25% αντίστοιχα relative to Expo-drop. This is a good tradeoff!

B. Simulation of scenarios with combined RT and NRT traffic:

SCENARIO #3 : We are evaluating the performance of Expo-drop, MLWDF, CD_EDD, CD_EDD_balanced, CD_EDD_mybalance, Expo-Linear, Expo-predict, Expo-weight, Expo-balanced, Expo-pred-wei and Expo-pred-balanced in a scenario with 8 VIDEO and 6 FTP users. The mean values of Es/No for each VIDEO user are: User1 : 5 dB, User2 : 8 dB, User3 : 10 dB, User 4: 13 dB, User5 : 16 dB, User6 : 19 dB, User 7 : 22 dB, User 8 : 25 dB, while for each FTP user are: User1 : 7 dB, User2 : 14 dB, User3 : 21 dB, User 4: 28 dB. Minimum required throughput is 64 Kbps for each FTP user.We also have: $T_i = 1\,\mathrm{sec}$ *και* $\delta_i = 10^{-2}$*, for FTP users and:* $T_i = 400m\,\mathrm{sec}$ *and* $\delta_i = 10^{-2}$ *for Video users.*

We note that what is important for video users is the ratio of packets being rejected due to deadline expiration. For this reason, in scenarios with combined RT and NRT traffic the performance criteria are: a) whether the NRT users get their minimum required throughput, b) the total cell throughput and c) the dropping rate of RT users.

Simulation results of scenario #3 are presented in Figures 7-8.

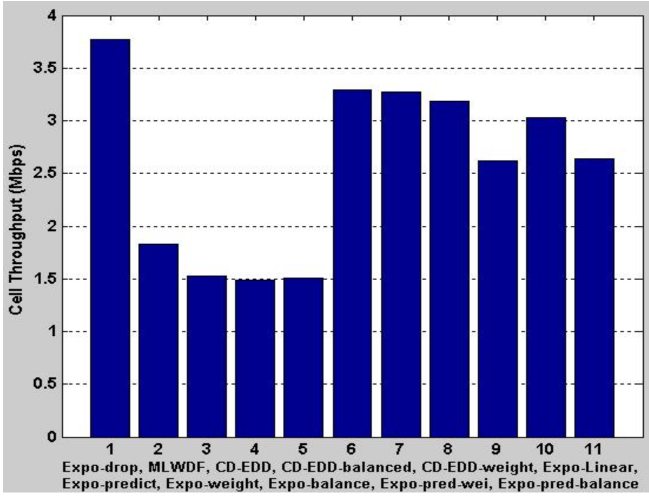

Figure 7 : Total cell throughput achieved by each algorithm.

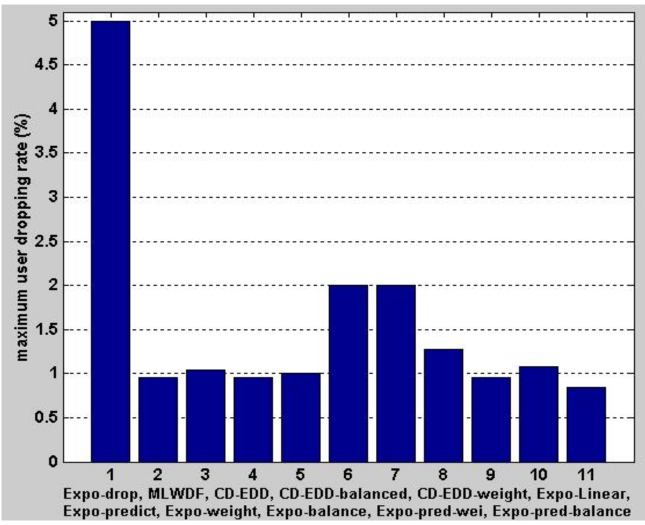

Figure 8 : Maximum dropping rate among RT users achieved by each algorithm..

MLWDF and CD_EDD achieve the same level of dropping rate as Expo-pred-wei does (Figure 8) but their cell throughput is about 50% lowe than that of Expo-pred-wei (Figure 7). At the same time cell throughput of Expo-drop is 25% higher than Expo-pred-wei's but its dropping rate is five times higher than Expo-pred-wei's.

SCENARIO #4 : We are evaluating the performance of Expo-drop, MLWDF, CD_EDD, CD_EDD_balanced, CD_EDD_mybalance, Expo-Linear, Expo-predict, Expo-weight, Expo-balanced, Expo-pred-wei and Expo-pred-balanced in a scenario with 10 VIDEO and 20 WWW users .The mean values of Es/No for each VIDEO user are: User1 : 8 dB, User2 : 10 dB, User3 : 12 dB, User 4: 14 dB, User5 : 16 dB, User6 : 18 dB, User7 : 20 dB, User8 : 22 dB, User9 : 24 dB, User10 : 28 dB, while for each WWW user are: User1 : 7 dB, User2 : 8 dB, User3 : 9 dB, User 4: 10 dB, User5 : 12 dB, User6 : 13 dB, User7 : 14 dB, User8 : 16 dB, User9 : 17 dB, User10 : 19 dB, User11 : 21 dB, User12 : 23 dB, User13 : 25 dB, User14 : 27 dB, User15 : 29 dB, User16 : 31 dB, User17 : 5 dB, User18 : 10 dB, User19 : 14 dB, User20 : 18 dB. We also have: $T_i = 4\,\mathbf{sec}$ *and* $\delta_i = 10^{-2}$ *for WWW users, and:* $T_i = 400m\,\mathbf{sec}$ *και* $\delta_i = 10^{-2}$ *for VIDEO users.*

Simulation results of scenario#4 are presented in Figures 9-10.

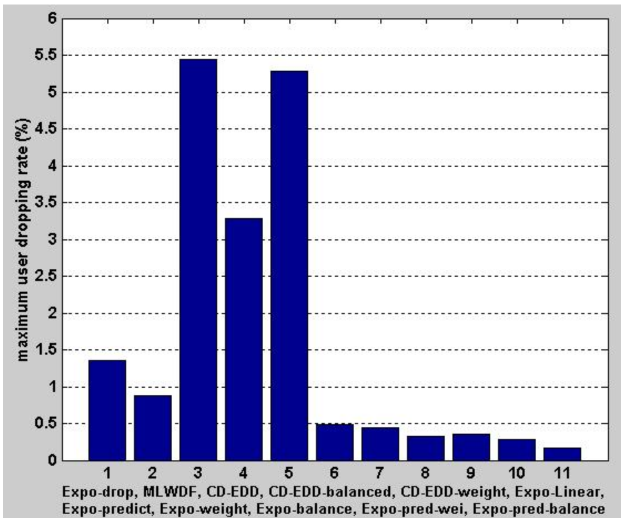

Figure 9 : Maximum dropping rate among VIDEO

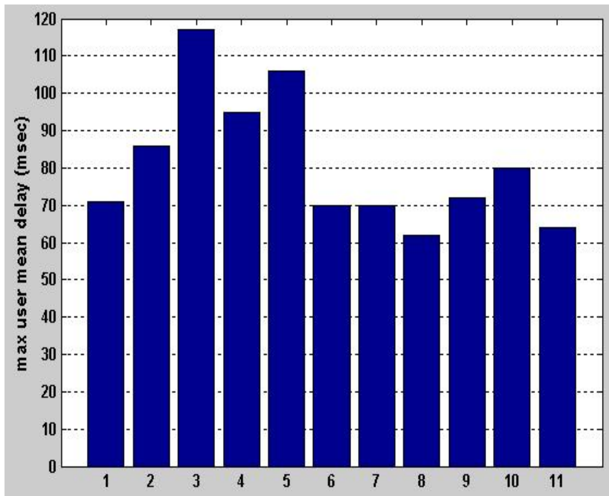

Figure 10 : The maximum value of mean delays of VIDEO users achieved by each algorithm.

All algorithms achieve good performance as far as delay of WWW users is concerned (Figure 10). Despite that, dropping rate achieved by Expo-pred-wei is 80% lower than that of Expo-drop's, 68% lower than MLWDF's and 92% lower than CD_EDD's (Figure 9).

C. Simulation of a RT traffic scenario

SCENARIO #5 : We are evaluating the performance of Expo-drop, MLWDF, CD_EDD, CD_EDD_balanced, CD_EDD_mybalance, Expo-Linear, Expo-predict, Expo-weight, Expo-balanced, Expo-pred-wei and Expo-pred-balanced in a scenario of 10 VIDEO users. The mean values of Es/No for each user are: User1 : 5 dB, User2 : 7 dB, User3

: 9 dB, User 4: 11 dB, User5 : 13 dB, User6 : 15 dB, User 7 : 17 dB, User 8 : 19 dB, User 9 : 21 dB, User 10 : 23 dB. We also have $T_i = 400m\sec$ *and* $\delta_i = 10^{-2}$*, the same for all users..*

Simulation results of scenario #5 are presented in Figures 11-12.

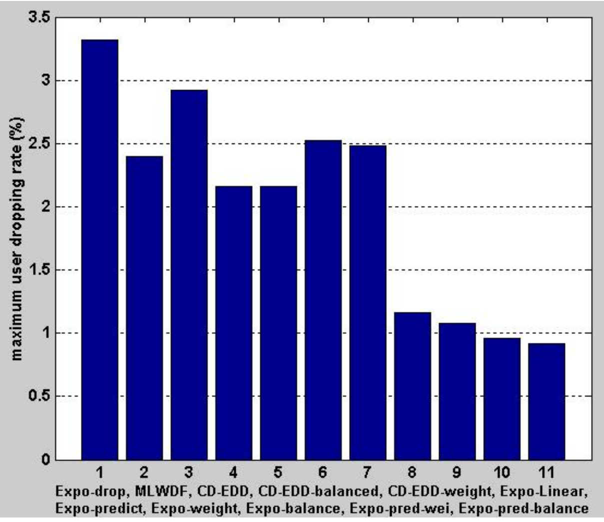

Figure 11 : Maximum dropping rate among VIDEO users achieved by each algorithm.

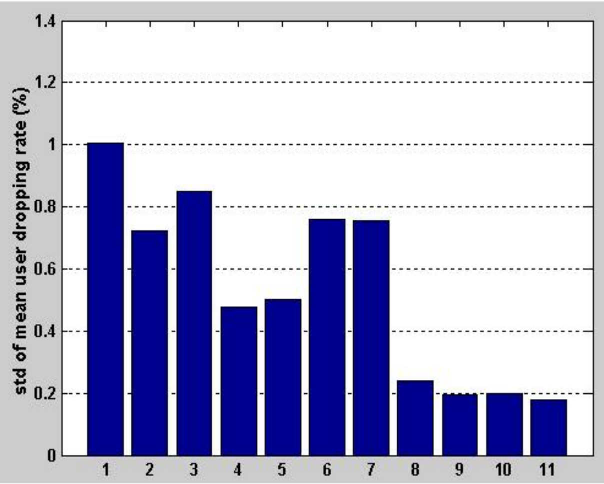

Figure 12 : Standard deviation of mean packet delay among VIDEO users, for each scheduling algorithm.

Expo-pred-wei's performance is among the best ones. CD_EDD and MLWDF have a medium performance, while Expo-drop has very bad performance. Expo-pred-wei's dropping rate is about 60% lower than CD_EDD's and MLWDF's and 70% lower than Expo-drop's (Figure11). As can be seen (Figure 12) balancing technique cooperates better with Expo-pred-wei than with the other algorithms.

We note here that the superior behaviour of Expo-pred-wei is mainly due to the modifications of Expo-Linear and Expo-weight. Expo-predict offers only a small improve over Expo-Linear's performance in scenario #4 and scenario #5.

VIII. CONCLUSIONS

In this paper we have proposed a new scheduling algorithm (Expo-pred-wei) for downlink shared multirate channels with varying condition and compared its performance to other existing algorithms under a wide range of traffic scenarios. Our final conclusion is that Expo-pred-wei succeeds in achieving the required QoS of each user quite well, while at the same time it achieves high levels of total throughput in all of the traffic scenarios examined. This goal is not achieved by the other algorithms examined. This fact makes Expo-pred-wei attractive for use regardless of the traffic types served by the system. This is a very desirable feature since cellular systems of 2.5G and 3.5G serve a wide variety of traffic whose mix changes over time, and therefore the performance of the scheduling algorithm should be independent of the traffic type served by the system.

REFERENCES

[1] Vadunur Bharghavan, Songwu Lu and Thyagarajan Nandagolap, "Fair Queueing in Wireless Networks: Issues and Approaches", IEEE Personal Communications, pp.44-53, February 1999
[2] Fragouli, Sivaraman and Srivastava , "Controlled Multimedia Wireless Link Sharing via Enhanced Class-based Queueing with Channel State Dependent Packet Scheduling, Proc. IEEE Infocom, April 1998
[3] Thyagarajan Nandagopal, Songwu Lu and Vaduvur Bharghavan, "A Unified Architecture for the Design and Evaluation of Wireless Fair Queueing Algorithms" , In Proceedings of ACM/MOBICOM 1999, August 1999
[4] "EDGE: Introduction of high-speed data in GSM/GPRS networks", White Paper, Ericsson, 2003 , [Online]. Available: **http://www.itpapers.com**
[5] Paul Bender, Peter Black, Mathew Grob, Roberto Padovani, Nagabhushana Sindhushayana and Andrew Viterbi, "CDMA/HDR: A Bandwidth-Efficient High-Speed Wireless Data Service for Nomadic Users", IEEE Communications, pp.70-77, July 2000
[6] 3GPP Technical Report TR 25.858 v5.0.0 (2002-03), High Speed Downlink Packet Access: Physical Layer Aspects (Release 5)
[7] Carl Eklund, Roger B. Marks, Kenneth L. Stanwood and Stanley Wang, "IEEE Standard 802.16: A Technical Overview of the WirelessMAN Air Interface for Broadband Wireless Access", IEEE Communications Magazine, June 2002
[8] Theodore Rappaport, "Wireless Communications : Principles and Practice", Prentice Hall, Second Edition 2002
[9] Michele Boaretto, Diego Angelini and Michele Zorzi, Fast Scheduling Policies using WCDMA High Speed Downlink Packet Access", IEEE Vehicular Technology Conference, October 2003
[10] T.J Mousley, "Performance of UMTS High Speed Downlink Packet Access for Data Streaming Applications", 3G Mobile Communication Technologies, Conference Publication No.489, May 2002
[11] Anders Foruskar, Stefan Parkvall, Magnus Persson and Maria Samuelsson, "Performance of WCDMA High Speed Packet Data", 2002 , [Online]. Available: **http://citeseer.ist.psu.edu/514957.html**
[12] Troels E. Kolding, "Link and System Performance Aspects of Proportional Fair Scheduling in WCDMA/HSDPA", Vehicular Technology Conference, VTC 2003 Fall
[13] Troels E. Kolding, Frank Frederiksen and Preben E.Morgensen, "Performance Aspects of WCDMA Systems with High Speed Downlink Packet Access (HSDPA)", IEEE Vehicular Technology Conference, September 2002

[14] Jonas Oberg, "Evaluation of Scheduling Algorithms for a Web Browsing Service in WCDMA High Speed Downlink Shared Channel", Master Thesis, Lulea University of Technology, Dep. of Computer Science and Electrical Engineering, 2002 , [Online]. Available:**http://epubl.luth.se/1402-1617/2002/126/LTU-EX-02126-SE.pdf**

[15] Yoshiaki Ofuji, Akihito Morimoto, Sadayuki Abeta and Mamoru Sawahashi, "Comparison of packet scheduling algorithms focusing on user throughput in high speed downlink packet", PIMRC 2002

[16] Matthew Andrews, Krishnan Kumaran, Kavita Ramanan, Alexander Stolyar, Rajiv Vijayakumar and Phil Whiting, "CDMA Data QoS Scheduling on the Forward Link Variable Channel Conditions", Bell Labs Technical Report, April. 2000

[17] Matthew Andrews, Krishnan Kumaran, Kavita Ramanan, Alexander Stolyar, Phil Whiting and Rajiv Vijayakumar, "Providing Quality of Service over a Shared Wireless Link", IEEE Communications Magazine, February 2001

[18] Sanjay Shakkotai and Alexander Stolyar, "Scheduling for Multiple Flows Sharing a Time-Varying Channel: The Exponential Rule", Bell Labs Technical Report, December 2000

[19] Ahmed K.F. Khattab and Khaled M.F. Elsayed, "Channel-Quality Earliest Deadline Due Fair Scheduling Schemes for Wireless Networks", ACM MSWiM, October 2004

[20] Sanjay Shakkotai and Alexander L. Stolyar, "Scheduling Algorithms for a Mixture of Real-Time and Non-Real-Time Data in HDR", Proc. of 17th International Teletraffic Congress, (ITC-17), pp.793-804, September 2001

[21] W.C. Jakes, Microwave Mobile Communications, Piscataway, NJ : IEEE Press, 1994

[22] A. L. Stolyar and K. Ramanan, "Largest Weighted Delay First Scheduling: Large Deviations and Optimality", Annals of Applied Probability, vol.11 (2001), No.1

TABLE I

Scheduling Scheme		**Case Description**		
	(explanation of arithmetic values)	Low QoS-cruciality with large deviations	High QoS cruciality with relatively low deviations	Same relative but different absolute QoS cruciality (the other values are the same or almost the same)
	Arithmetic values taken by the parameters of the scheduling criterion	$\delta_1 = 0.01$, $\delta_2 = 0.01$ $T_1 = 0.05$ sec $T_2 = 0.05$ sec $R_1 = 360$ Kbps $\bar{R}_1 = 360$ Kbps $R_2 = 740$ Kbps $\bar{R}_2 = 1320$ Kbps $D_1 = 0.005$ sec $D_2 = 0.01$ sec	$\delta_1 = 0.01$, $\delta_2 = 0.01$ $T_1 = 0.05$ sec $T_2 = 0.05$ sec $R_1 = 360$ Kbps $\bar{R}_1 = 360$ Kbps $R_2 = 740$ Kbps $\bar{R}_2 = 1320$ Kbps $D_1 = 0.032$ sec $D_2 = 0.048$ sec	$\delta_1 = 0.01$, $\delta_2 = 0.01$ $T_1 = 0.05$ sec $T_2 = 0.1$ sec $R_1 = 360$ Kbps $\bar{R}_1 = 360$ Kbps $R_2 = 1440$ Kbps $\bar{R}_2 = 1320$ Kbps $D_1 = 0.04$ sec $D_2 = 0.08$ sec
M-LWDF	Priority of user 1	0.2	1.28	1.6
	Priority of user 2	0.22	1.08	1.74
CD-EDD	Priority of user 1	4.4	71	160
	Priority of user 2	5.6	538	87
Expo-Full	Priority of user 1	37	35	40
	Priority of user 2	24	26	21.8
Expo-drop	Priority of user 1	45	70	81
	Priority of user 2	29	52	44
Expo-Linear	Priority of user 1	49	144	198
	Priority of user 2	33	153	108
		Who should be scheduled: User-1 Who is scheduled: M-LWDF: User-2 CD-EDD: User-2 **Expo-Full: User-1** **Expo-drop: User-1** **Expo-Linear: User-1**	Who should be scheduled: User-2 Who is scheduled: M-LWDF: User-1 **CD-EDD: User-2** Expo-Full: User-1 Expo-drop: User-1 **Expo-Linear: User-2**	Who should be scheduled: User-1 Who is scheduled: M-LWDF: User-2 **CD-EDD: User-1** **Expo-Full: User-1** **Expo-drop: User-1** **Expo-Linear: User-1**