An Opportunist extension of Wireless Fair Service for Packet Scheduling in OFDMA

Carle Lengoumbi, Philippe Martins, Philippe Godlewski
Telecom Paris - GET
LTCI-UMR 5141 CNRS
46 rue Barrault 75013 Paris, France
Email:{carle.lengoumbi, philippe.martins, philippe.godlewski}@enst.fr

Abstract— **In this paper, we investigate packet scheduling and radio resource allocation in Orthogonal Frequency Division Multiple Access (OFDMA) wireless networks. Each subchannel is a Rayleigh fading channel modeled by a Finite State Markov Channel (FSMC). The Wireless Fair Service (WFS) packet Scheduler is modified for a multi carrier and multi rate context. The proposed algorithm, named Opportunist Wireless Fair Service (OWFS), copes with both real time (RT) and non real time (NRT) traffic constraints. For both traffic classes, specific criteria of performance are compared with PLFS, a scheduler recently proposed for OFDMA.**

Index Terms: Scheduling, Resource Allocation, Opportunist Wireless Fair Service, OFDMA.

I. INTRODUCTION

Orthogonal Frequency Division Multiple Access (OFDMA) is a promising multiple access technique for next wireless broadband networks. By dividing the bandwidth into multiple subsets of subcarriers (named subchannels), OFDMA exploits the multi-user diversity and thus increases the system throughput. In addition, Adaptive modulation and coding (AMC) can be employed. AMC has been recognized as a key technique to enhance the resource utilization efficiency and provide better QoS (Quality of Service) guarantees.

Over the last years, resource allocation and scheduling in OFDMA has been subject to active research. Several papers ([1]-[3]) focus on resource allocation from the physical layer point of view. In [1]-[2], the cell capacity is maximized subject to a power constraint and minimum rate requirements. Authors of [3] consider proportional rate constraints during cell capacity optimization. Scheduling algorithms ([4]-[8]) deal with multiplexing at a packet level. Unlike assumptions in [1]-[3], buffers have a finite length and may be empty from time to time; QoS of multimedia traffic is more evolved than a user rate requirement. Some papers ([4]-[6]) concentrate on one single type of traffic. In [4], the metric proposed for real time (RT) traffic is a function of the current and maximum packet loss rate of the flow and its head of line (HoL) packet delay. In [5], a generalization of the conventional proportional fair (PF) scheduler is proposed for non real time (NRT) traffic. Weighting factors are introduced on both instantaneous and average data rate. On each subchannel, the flow maximizing the metric can transmit. In [6], three scheduling schemes based on the PF scheduler are compared. Regarding the global throughput, the authors show that one global scheduler outperforms the scheme consisting in one independent scheduler per subchannel.

In this paper, we want to satisfy simultaneously QoS requirements of data rate-sensitive and delay-sensitive traffics. We also want a global scheduler to handle all the subchannels. Wireless Fair Service (WFS) has been introduced ([5]) to adapt fair packet scheduling to wireless networks. WFS introduces a delay weight in addition to the rate weight already existing in Weigthed Fair Queuing (WFQ); WFQ is a wire-line fair queuing algorithm that allocates resources based on the General Processor Sharing (GPS) model. WFS supports QoS of both error and delay sensitive flows but for one single rate channel. We are interested in adapting WFS to OFDMA. The modified scheduler copes with multi-carrier and multi-rate aspects; it is called Opportunistic Wireless Fair Service (OWFS). The virtual start tag and virtual finish tag are modified to take account of users' channel states. Once selected, a flow receives his best available subchannel. The goal of the paper is to compare this opportunist "GPS based" algorithm with a "PF based" algorithm. In the last mentioned class of algorithms, we choose the Packet Loss Fair Scheduling (PLFS) scheduler proposed in [8]. As [4]-[6], a "PF based" metric is used to prioritize flows and it considers both RT and NRT traffic classes.

In this paper, downlink transmissions of a single cell are considered. Subchannels consist of adjacent subcarriers within the channel coherence bandwidth. To cope with fast and slow channel variations, AMC is used as link adaptation technique. A subchannel is modeled as a Finite State Markov Chain (FSMC) to handle both frequency and time channel variations. Each state in a FSMC represents one AMC mode. Users feed back their CSI (Channel State Information) as a per subchannel AMC mode. In the sequel, each user is associated with only one flow. The remainder of the paper is organized as follows. The next section describes the Wireless Fair Service algorithm. Section III presents the OWFS packet scheduler. In section IV, OWFS performance is compared with the PLFS ([8]). Global throughput, average delay for both traffic classes and RT packet drop rate are considered. Section V concludes the paper.

II. WIRELESS FAIR SERVICE (WFS)

In this section, the main features of WFS described in [7] are outlined. WFS has been designed for wireless error prone links with a simple on-off channel error model. In section III, OWFS is defined to overcome this main limitation of WFS.

A. Error free service model

In the wireless context, the "error-free service model" deals with the amount of service allocated per flow when there are no channel errors. WFS error-free service model is an extension from WFQ. Each flow is assigned a weight for rate r_i and delay Φ_i. The delay weight avoids the coupling between the delay observed by packets and the fraction of bandwidth given to the flow (section 4.2 in [7]). The k^{th} arriving packet of flow i causes the computation of a virtual start tag $S(p^k_i)$ and a virtual finish tag $F(p^k_i)$:

$$S(p^k_i) = \max(V(A(p^k_i)), S(p_i^{k-1}) + L_i^{k-1}/r_i)$$

$$F(p^k_i) = S(p^k_i) + L_i^k/\Phi_i$$

where $V(A(p^k_i))$, is the virtual time when packet p^k_i arrives and L_i^k is the length of the k^{th} packet of flow i. The virtual time $V(t)$ is computed from the set of backlogged flows $B(t)$ (i.e. flows that have packets to send) of the error-free service: $dV(t)/dt = C / \Sigma_{i\in B(t)} r_i$ where C is the total capacity of the server and r_i is the rate weight of an individual flow i. The flow with the lowest service tag can transmit. Next subsection defines the WFS service tag.

B. Slot queue and Packet queue

A flow has one packet queue and one "slot" queue. A slot is a sort of transmission opportunity. At any time, the number of slots in the slot queue is the same than the number of packets in the packet queue. Each time a packet is added to the packet queue, a new slot is added to the slot queue. The corresponding start and finish tags are associated with the slot, not with the packet. The service tag of a flow is the finish tag of the head of line (HoL) slot. The flow selected for transmission abides by two conditions. The first condition (denoted *condition1*) states that the flow must have a HoL virtual start tag lower than the current virtual time plus a lookahead parameter ρ. The second condition (denoted *condition2*) states that the flow must have the lowest service tag (among flows satisfying *condition1*). The selected flow transmits the HoL packet of the packet queue and deletes the HoL slot of the slot queue.

The slot queue allows decoupling of connection level packet management policies from link level packet scheduling policies ([7]). Both delay sensitive and error sensitive flows can be handled. If a HoL packet is dropped (for exceeding the maximum number of retransmissions or a delay bound), the HoL slot should be unchanged. Thereby, the internal packet management (queuing or dropping) policy does not modify the transmission opportunities of a flow.

C. Compensation model

A complete description of the compensation model lies in [7]. Only key points are given here. If the error free service model selects a flow perceiving a bad channel, the compensation model decides which flow can transmit. If another flow can transmit instead, the initial flow receives compensation later. A flow may be leading, lagging or *in sync* as it transmits more, less or the same amount of packets than the corresponding flow that always perceives a clean channel. Each flow has counters to measure its lead or its lag. A leading flow relinquishes a proportion $E(i)/E_{max}(i)$ of its transmission opportunities, where $E(i)$ is the leading counter and $E_{max}(i)$ is the leading bound. If a transmission opportunity is relinquished by a leading flow i or if the initial flow i has a bad channel, the compensation model proceeds in a predefined order to find another flow. It looks through the set of backlogged lagging flows, then through the set of backlogged leading flows ($E(j) < E_{max}(j)$) and in third position, through the set of backlogged *in sync* flows. If a flow j is found with a good channel, it transmits its HoL packet and counters (of flows i and j) are updated. Otherwise any backlogged flow j with $E(j) = E_{max}(j)$ and which has a good channel can transmits its HoL packet. If another flow j transmits, the initial flow i leaves the packet queue unchanged but should delete the HoL slot (since flow i gave up the transmission opportunity and should not capture the entire channel as soon as it has a good channel). In that case, flow i can create a slot at the end of the slot queue to keep equal length between slot and packet queues.

III. SCHEDULERS FOR OFDMA

A. Channel model for OFDMA

At each scheduling period, S subchannels are available for allocation. We consider the downlink of one single cell. Subchannels are independent from each other. Due to Raleigh fading, a user experiments different signal to noise ratios (SNR) on the distinct subchannels. Perfect CSI knowledge is assumed. The SNR range is partitioned and mode m corresponds to the SNR interval $[\gamma_m, \gamma_{m+1}[$. A different modulation and coding scheme (MCS) is used on each mode (cf. Table1). Each subchannel is modeled according to a FSMC Channel Model. The adjacent state transition probabilities between the different modes are determined as in [9]. For one user, the SNR variations in time, on a subchannel, are thus determined by the transition matrix.

To allocate the different subchannels among backlogged flows, we propose an enhancement of WFS. OWFS is a modified version of WFS that takes into account the instantaneous channel states of users during the scheduling.

B. Opportunistic Wireless Fair Service (OWFS)

In [10], the Weighted Fair Queuing (WFQ) scheduler is modified into Opportunistic Weighted Fair Queuing (OWFQ) to deal with a multi-rate channel in a TDMA context. The virtual start stag is unchanged. The new definition of the virtual finish tag includes the channel state $x_i(T)$ of flow i. The finish tag is only computed for HoL packets at the scheduling instant T:

$$F(p_i^{HOL}) = S(p_i^{HOL}) + L_i^{HOL} / (r_i\, x_i(T)).$$

Recall that in WFQ, $F(p^k_i) = S(p^k_i) + L_i^k / r_i$ for any k^{th} arriving packet of flow i; $S(p^k_i)$ is the start tag, L_i^k is the length of the packet and r_i is the weight of the flow i.

In WFS, there are two conditions to be scheduled: we saw that *condition1* involves the virtual start tag while *condition2* concerns the virtual finish tag. Adapting WFS to a multi-rate context thus requires a new definition for both virtual start tag and virtual finish tag.

TABLE I. TRANSMISSION MODES

Mode	SNR required	Packet per subchannel	Modulation	Coding rate
1	2.5 dB	1	BPSK	1/2
2	6 dB	2	QPSK	1/2
3	9.5 dB	3	QPSK	3/4
4	16 dB	6	16QAM	3/4
5	22.5 dB	9	64QAM	3/4

The new scheduler OWFS is built to take account of a multi-rate context. So, we extend the idea of [10] to definitions of both virtual start tag and virtual finish tag. In [10], there is no ambiguity on the channel state $x_i(T)$ because one single channel is considered. OWFS is built for OFDMA; we distinguish two channel modes: $m_{i,max}(t)$ and $m_{i,best}(t)$ to handle multiple multi-rate channels. The highest channel mode of flow *i* over all the subchannels is $m_{i,max}(t)$ while $m_{i,best}(t)$ is the highest channel mode of flow *i* over all the available subchannels. This distinction is helpful for new definitions of virtual start tag and virtual finish tag.

1) Expression of the virtual start tag

The k^{th} arriving packet of flow *i* causes the computation of a new virtual start tag $S(p^k_i)$:

$$S(p^k_i) = \max\left(V(A(p^k_i))\ ,\ S(p_i^{k-1}) + L_i^{k-1} / (r_i\ m_{i,max}(A(p^k_i)) \right)$$

When a packet p^k_i arrives at time $A(p^k_i)$, all the subchannels are available for the next scheduling instant. That explains why the channel state $m_{i,max}(t)$ is chosen. The higher $m_{i,max}(A(p^k_i))$, the lower $S(p^k_i)$; the more likely flow *i* respects *condition1*.

2) Virtual time in a multicarrier context

The virtual time $V(t)$ is computed from the set of backlogged flows $B(t)$: $dV(t)/dt = C_{avg} / \Sigma_{i\in B(t)}\ r_i$ where $C_{avg} = (\Sigma_{i\in B(t)} \Sigma_{s=1..S}\ C_{i,s}) / |B(t)|$. As $C_{i,s}$ is the capacity of the flow *i* on subchannel *s*, C_{avg} is the average global capacity seen by a backlogged flow.

3) Expression of the virtual finish tag

Among flows whose HoL virtual start tag satisfies *condition1*, the scheduler selects the flow with the minimum service tag until all subchannels are allocated. The service tag is the virtual finish tag of the HoL slot:

$$F(p_i^{HOL}) = S(p_i^{HOL}) + L_i^{HOL} / (\Phi_i\ m_{i,best}(t)).$$

4) Opportunist resource allocation principle

Each time a flow *i* is scheduled, it receives its "best available subchannel" (i.e. subchannel *s* such that $m_{i,s} = m_{i,best}(t)$). The set of available subchannels steps down. That explains why $m_{i,best}(t)$ is used rather than $m_{i,max}(t)$ in the virtual finish tag computation. If the initial flow *i* cannot transmit on its "best available subchannel" (i.e. $m_{i,best}(t)$=0) the compensation model is launched. The compensation model is nearly unchanged compared to WFS. Nevertheless, between several backlogged flows from the same category (lagging, leading or sync flows), the one with the "best available subchannel" is preferred. If at the end of the compensation block (cf. Fig.1), no flow has been found, the remaining available channels are lost. Indeed, in that case, any backlogged flows *j* sees an available channel *s* with the AMC mode $m_{j,s}$=0.

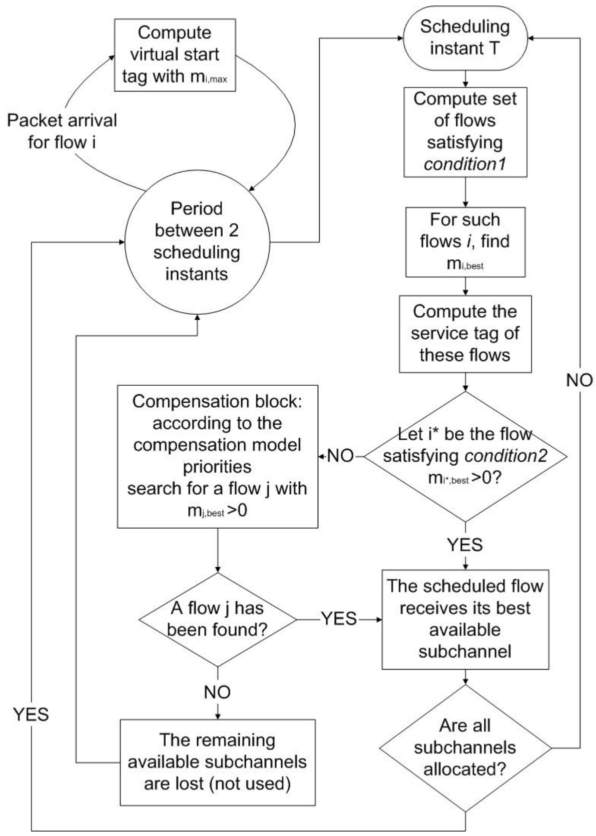

Figure 1. OWFS scheduling procedure

C. Packet Loss Fair Scheduling (PLFS)

For comparison purpose, we are interested in the PLFS proposed in [8]. The priority metric used to schedule RT traffic differs from that used for NRT traffic. Each time a flow *i* is scheduled, it receives its best available subchannel. A proportion α of the bandwidth is reserved to RT traffic. This bandwidth reservation during one slot is an adaptation of the RT and NRT regions during one frame in [8].

Until the reserved proportion of subchannels is allocated and as long as there are backlogged RT flows, the scheduler selects the RT flow with highest priority p_{RT} : $p_{RT}(i) = (m_{i,best}(t) / M_i) . PDR_i(t) / (PDR_{req,i} . D_{max,i})$; where $m_{i,best}(t)$ is the highest MCS level of flow *i* among remaining subchannels, M_i is the average MCS level of flow *i* on a sliding window, *PDR* stands for Packet Drop Rate (a flow *i* packet whose delay is beyond $D_{max,i}$ is dropped).

After scheduling of RT traffic, there are at least $(1-\alpha)S$ subchannels left. A distinct rule p_{NRT} is used for NRT traffic: $p_{NRT}(i) = L_i(t)\ D_i(t)\ m_{i,best}(t) / M_i$; where $L_i(t)$ is the queue length of flow *i* and $D_i(t)$ is the HoL packet delay of flow *i*.

In PLFS, the average channel state M_i of flow *i* is computed from a sliding window (M_i may be taken on subchannel *s* such that $m_{i,s} = m_{i,best}(t)$) while in PF, the average channel state is considered up to time *t*. However, from the ratio between an instantaneous channel state and an averaged channel state, the PLFS scheduler can be classified into the class of "PF based" scheduling algorithms.

IV. SIMULATION PARAMETERS AND RESULTS

A. *Physical layer assumptions*

We consider an OFDMA system with 768 data subcarriers grouped in S = 16 subchannels. Two subcarriers are spaced out around 10 kHz. Remaining parameters are listed in Table II. Subchannels are allocated with equal power at the beginning of every frame. The frame duration is 5 ms. Each simulation had a typical run of 24000 frames; the simulation time is thus 2 minutes. Each result is averaged over 25 simulation runs.

B. *Traffic model*

Two traffic classes (RT and NRT) are considered for the traffic input. We have considered a Poisson arrival process and a bursty traffic. Plots corresponding to a Poisson arrival process are omitted due to lack of space. Regarding the bursty traffic, both RT and NRT are generated by an ON-OFF model. For RT traffic, ON and OFF periods have an exponential distributed duration. Mean duration are respectively 1 sec and 1.35 sec. RT packets arrive at a constant rate of 64 kbps in the ON state. As NRT traffic, we assume FTP sessions. The duration of the ON state depends on the file size; the file size distribution is truncated lognormal (mean: 1 Mbytes, max: 2 Mbytes, std: 0.3 Mbytes). NRT packets arrive at a constant rate of 384 kbps during the ON state. The OFF state corresponds to the reading time; the duration is exponentially distributed with a mean of 15 sec.

C. *Schedulers parameters*

In our simulation, we assume that fixed length packets are stored into the scheduler buffers. RT packets are dropped if their queuing time exceeds D_{max}=100ms.

1) *PLFS parameters*

The maximum packet drop rate PDR_{req} is set to 10^{-3}. The percentage of bandwidth reserved to backlogged RT traffic is 60% (α=0.6).

2) *OWFS parameters*

Regarding OWFS, rate weights are set as 1 for both RT and NRT. Because RT traffic has a deadline to observe, the RT delay weight is set as 40 while it is set as 1 for NRT traffic. The lookahead parameter is set as $\rho=\infty$.

TABLE II. SIMULATION PARAMETERS

Parameters	Value
Number of subcarriers	1024
Number of data subcarriers	768
Number of subchannels	16
Packet (or MAC PDU) size	60 bytes
Maximum Packet Drop Rate	10^{-3}
Maximum packet delay	100 ms
Central frequency	2 GHz
Bandwitdth	10 MHz
BS transmission power	43 dBm
Cell radius	2 km
User distribution	uniform
Path Loss Model	Cost Hata

D. *Schedulers performance*

We compare in the sequel, the performances of PLFS and OWFS schedulers with respect to global throughput, mean delay and RT drop rate. The total number U of users varies from 20 to 120. In each case, 80% of users are dedicated to RT traffic and 20% to NRT traffic.

1) *Global throughput*

The average throughput of the cell is plotted in Fig.2 according to the number of users. OWFS NRT throughput outperforms PLFS. The same behaviour is observed for global throughput because OWFS and PLFS RT throughput are nearly the same. At heavy loads, the NRT throughput is bounded in PLFS because of RT bandwidth reservation.

2) *Delay and Drop Rate of RT traffic*

Fig. 3 shows the average delay of transmitted packets for RT traffic. It can be seen that OWFS RT packets have a lower average delay than PLFS RT packets. The maximum delay of a PLFS RT transmitted packet is very close to the delay bound D_{max} (for U=60 and beyond); the PLFS drop rate is thus expected to be higher than OWFS (since a RT packet is dropped beyond D_{max}). Fig.5 confirms that the packet drop rate (PDR) of OWFS is better than that of PLFS.

In Fig.4, for $0 \leq U \leq 100$, the delay experienced by PLFS transmitted packets suffers from greater variations (jitter) than that of OWFS. This can be challenging for real time traffic such as voice and video. In this case, PLFS is more convenient for streaming.

3) *Delay of NRT traffic*

Fig.6 shows that OWFS NRT packets have a lower average delay than PLFS NRT packets. For both schedulers, NRT packets average delay is significantly higher than RT packets average delay. The reasons are simple. In PLFS, a greater bandwidth is reserved to RT traffic. In OWFS, RT traffic has a greater delay weight than NRT traffic.

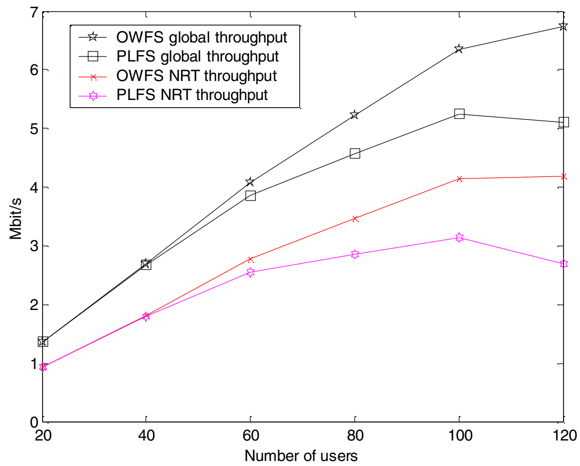

Figure 2. Throughput according to the number of users

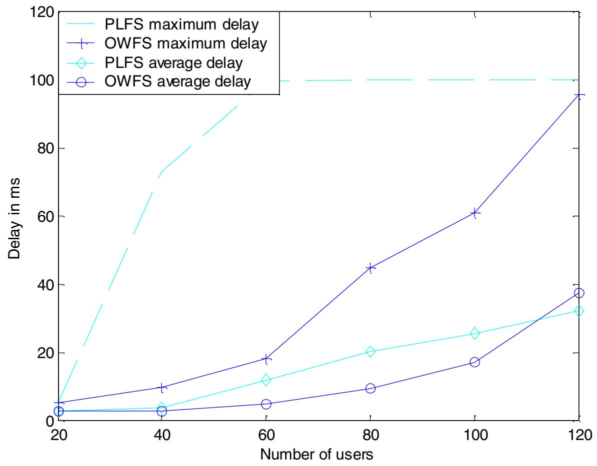

Figure 3. Average and maximum delay for RT packets

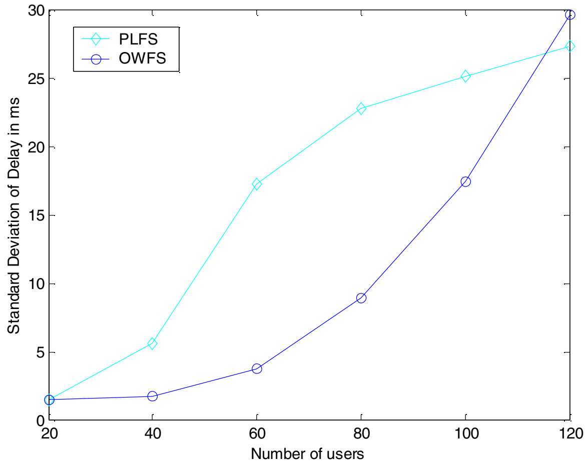

Figure 4. Standard Deviation of the delay for RT packets

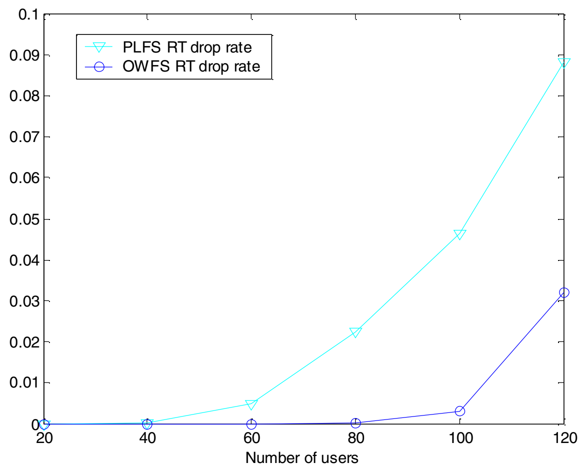

Figure 5. Packet drop rate of RT traffic

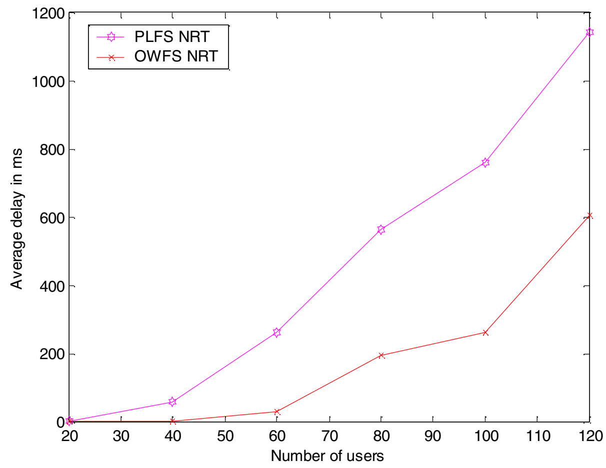

Figure 6. Average delay of NRT traffic

V. CONCLUSION

We have studied an Opportunistic Wireless Fair Service (OWFS) packet scheduler in the OFDMA context. OWFS is able to deal with different QoS classes without distinct priority rules or predefined bandwidth reservation for RT traffic. From comparison with PLFS, a recently proposed scheduler, it has been seen that OWFS performs well in the OFDMA context regarding average delays and global throughput. Regarding RT traffic, OWFS shows small delay variations for moderate loads. Ongoing work seeks to provide an analytical characterization of OWFS, especially the schedulable region according to the load.

VI. REFERENCES

[1] H. Yin, H. Liu, “An Efficient multiuser Loading Algorithm for OFDM-based Broadband Wireless Systems”, IEEE Global Telecommunications Conference, no. 1, Nov 2000, pp. 103– 107.

[2] J. Gross, H. Karl, F. Fitzek, A. Wolisz, “Comparison of heuristic and optimal subcarrier assignment algorithms”, IEEE International Conference on Wireless Networks (ICWN) 2003, pp. 249-255.

[3] I. C. Wong, Z. Shen, J.G.Andrews, B.L.Evans, “A Low Complexity Algorithm for proportional Resource Allocation in OFDMA Systems”, IEEE Signal Processing Systems, SIPS 2004. Page(s):1 – 6.

[4] S. Shin, B. Ryu. “Packet loss fair scheduling scheme for real-time traffic in ofdma systems”. ETRI Journal, vol.26, no.5, Oct. 2004, pp.391-396.

[5] C. Wengerter, J. Ohlhorst, A. Golitschek, E. von Elbwart. “Fairness and Throughput Analysis for Generalized Proportional Fair Frequency Scheduling in OFDMA” VTC spring 2005.

[6] W. Anchun, X. Liang, Z. Shidong, X. Xibin, Y. Yan. “Dynamic resource management in the fourth generation wireless systems”**.** ICCT. Date: 9-11 April 2003.On page(s): 1095- 1098 vol.2

[7] S. Lu, T. Nandagopal, V. Bharghavan, “A Wireless Fair Service Algorithm for Packet Cellular Networks”. MOBICOM '98, Oct. 25-30.

[8] S. Shin, S. Bahng, I. Koo and K. Kim. “QoS-Oriented Packet Scheduling Schemes for Multimedia Traffics in OFDMA Systems”. Networking - ICN 2005 Volume 3420/2005 on pages 153-160.

[9] Q. Liu, S. Zhou, G. B. Giannakis, “Cross-Layer Modeling of Adaptive Wireless Links for QoS Support in Heterogeneous Wired-Wireless Networks”. Wireless Networks, Volume 12, Issue 4, Pages: 427 - 437.

[10] K. Khawam, “Opportunistic Weigthed Fair Queuing”. VTC fall 2006, September 25-28, Montreal, Canada.