Compensated Proportional Fair Scheduling in Multiuser OFDM Wireless Networks

Cédric Gueguen and Sébastien Baey
UPMC Univ. Paris 06
104, avenue du Président Kennedy, 75016 Paris, France
Email: {cedric.gueguen, sebastien.baey}@lip6.fr

Abstract—In wireless networks, providing a fair bandwidth allocation without too much reducing the system throughput is very challenging. Path loss attenuations induce unequal spectral efficiencies and in terms unequal throughputs for mobiles with different geographical positions. In the literature, Proportional Fair (PF) is acknowledged as the reference scheduler in multiuser OFDM wireless networks. Opportunistically considering the channel state, PF is adapted to the wireless environment fighting the multipath fading negative effects. PF takes advantage of multiuser diversity and globally maximizes the throughput. Additionally, PF scheduling currently makes the best tradeoff between fairness and throughput maximization. However, severe fairness deficiencies appear when the mobiles experience unequal path loss. In this paper, we propose to solve this fairness issue with a modified PF scheme that introduces distance compensation factors. Simulation results show that this well-balanced resource allocation outperforms other existing scheduling schemes and jointly provides both high system throughput and high fairness.

Index Terms—Orthogonal Frequency Division Multiplexing, path loss, multipath fading, multiuser diversity, opportunistic scheduling, multimedia, Medium Access Control.

I. INTRODUCTION

Bandwidth allocation in next generation broadband wireless networks (4G systems) is a difficult issue. New multimedia services with tight QoS constraints require increased system capacity together with high fairness. Already implemented in 802.11a/g and 802.16, Orthogonal Frequency Division Multiplexing (OFDM) appears as the most promising physical layer technique for broadband transmission for its capacity to efficiently reduce the harmful effects of multipath fading in wireless transmissions [1]. Intensive research efforts have recently been given in order to propose efficient schedulers for OFDM based networks.

A. Classical Scheduling Algorithms

Round Robin (RR) and Maximum SNR1 (MaxSNR) [2], [3], [4] are two well-attested bandwidth allocation strategies in wireless networks. RR fairly allocates an equal share of the bandwidth to each mobile in a ring fashion. MaxSNR exploits the concept of opportunistic scheduling. Priority is given to the mobile which currently has the greatest signal-to-noise ratio (SNR). Profiting of the multiuser diversity and continuously allocating the radio resource to the mobile with the best spectral efficiency, MaxSNR strongly improves the

1also known as Maximum Carrier to Interference ratio (MaxC/I).

system throughput. However, a negative side effect of this strategy is that the closest mobiles to the access point have disproportionate priorities over mobiles more distant since their path loss attenuation is much smaller. This results in a severe lack of fairness as illustrated in Fig. 1.

B. The Proportional Fair Scheduler

The Proportional Fair (PF) scheduling scheme [5], [6] allocates units of bandwidth to the mobile which currently has the greatest SNR value with respect to its average experienced SNR. At a short time scale, path loss variations are negligible and channel state variations are mainly due to multipath fading, statistically similar for all mobiles. Thus, PF provides an equal sharing of the total available bandwidth among the mobiles as RR. Applying the opportunistic scheduling technique, system throughput maximization is also obtained as with MaxSNR. PF actually combines the advantages of the classical schemes and currently appears as the best bandwidth management scheme. However, severe fairness deficiencies remain when considering mobiles with unequal spatial positions, as shown in Fig. 1. Indeed, since the farther mobiles have a lower spectral efficiency than the closer ones due to pathloss, all mobiles do not all benefit of an equal average throughput despite they all obtain an equal share of bandwidth. This induces heterogeneous delays and unequal QoS [7], [8].

C. The Compensated Proportional Fair Scheduler

With the objective of solving the PF fairness issues without sacrificing system capacity, this paper proposes a modified PF
scheduling scheme called “Compensated Proportional Fair” (CPF)\(^2\). CPF introduces correction factors in the PF in order to compensate the path loss negative effect on fairness while keeping the PF system throughput maximization properties. With this compensation, CPF is aware of the path loss disastrous effect on fairness and adequate priorities between the mobiles are always adjusted in order to ensure them an equal throughput. Additionally, this scheduling finely and simultaneously manages all mobiles and keeps a maximum number of them active across time with relatively low traffic backlogs. Preserving the multiuser diversity, CPF takes a maximal benefit of the opportunistic scheduling technique and maximizes the system capacity better than PF. Efficient support of multimedia services is provided well-combining the system capacity maximization and fairness objectives required for 4G OFDM wireless networks.

The outline of the paper is as follows. Section II gives a description of the system under study. Section III presents the scheduling algorithm. Performances are studied in Section IV. Section V concludes the paper.

**II. SYSTEM DESCRIPTION**

The physical layer operates using the structure described in Fig. 2 which ensures a good compatibility with the OFDM based transmission mode of the IEEE 802.16-2004 [9], [10]. The total available bandwidth is divided in sub-frequency bands (subcarriers). The radio resource is further divided in the time domain in frames. In the following, the frame duration is fixed to 2 ms. Each frame is itself divided in time slots. The time slot duration is an integer multiple of the OFDM symbol duration. The number of subcarriers is chosen so that the width of each sub-frequency band is less than the coherence bandwidth of the channel. Moreover, the frame duration is fixed to a value much smaller than the coherence time of the channel. With these assumptions, the transmission on each subcarrier is subject to flat fading with a channel state that can be considered static during each frame.

The elementary resource unit is defined as any (subcarrier, time slot) pair. Each of these resource units may be allocated to any mobile with a specific modulation order. Transmissions performed on different RUs by different mobiles are assumed to have independent channel state variations. On each RU, the modulation scheme is QAM with a modulation order adapted to the channel state between the access point and the mobile to which it is allocated. This provides the flexible resource allocation framework required for opportunistic scheduling.

The system is operated using time division duplexing with four subframes: the *downlink feedback subframe*, the *downlink data subframe*, the *uplink contention subframe* and the *uplink data subframe*. The uplink and downlink data subframes are used for the transmissions of user data. In the feedback and contention subframes, control information is communicated between the mobiles and the access point. This frame structure supposes a perfect time and frequency synchronization between the mobiles and the access point as described in [11]. Therefore, each frame starts with a preamble used for synchronization purposes.

This paper focuses on the proper allocation of radio resources among the set of mobiles situated in the coverage zone of an access point both in the uplink and in the downlink. The scheduling is performed during the uplink data transmission phase in a centralized approach. The packets originating from the backhaul network are buffered in the access point which schedules the downlink transmissions. In the uplink, the mobiles signal their traffic backlog in the contention subframe to the access point which builds the uplink resource mapping. The scheduler, located in the access point, grants resource units

\(^2\)This work has been partially supported by the IST European project WIP under contract 27402.
to each mobile as a function of:
- its data integrity requirements specified by a Bit Error Rate (BER) target which we denote by $BER_{target}$,
- its traffic backlog,
- its channel state.

Perfect knowledge of the channel state is assumed to be available at the receiver. The current channel attenuation on each subcarrier and for each mobile is estimated by the access point based on the SNR of the signal sent by each mobile during the uplink contention subframe. Assuming that the channel state is stable on a scale of 50 ms [12], the mobiles shall transmit their control information alternatively on each subcarrier so that the access point may refresh the channel state information once every 25 frames.

### III. The CPF Scheduling

At each scheduling epoch, the scheduler computes the maximum number of bits $B_{k,n}$ that can be transmitted in a time slot of subcarrier $n$ if assigned to mobile $k$, for all $k$ and all $n$. This number of bits is limited by two main factors: the data integrity requirement and the supported modulation orders.

The bit error probability is upper bounded by the symbol error probability [3] and the time slot duration is assumed equal to the duration $T_s$ of an OFDM symbol. The required received power $P_r(q)$ for transmitting $q$ bits in a resource unit while keeping below the data integrity requirement $BER_{target}$ is a function of the modulation type, its order and the single-sided power spectral density of noise $N_0$. For QAM and a modulation order $M$ on a flat fading channel [1]:

$$P_r(q) = \frac{2N_0}{3T_s} \left( erfc^{-1} \left( \frac{BER_{target}}{2} \right) \right)^2 (M - 1), \quad (1)$$

where $M = 2^q$ and $erfc$ is the complementary error function. $P_r(q)$ may also be determined in practice based on BER history and updated according to information collected on experienced BER.

The transmit power $P_{k,n}$ of mobile $k$ on subcarrier $n$ is upper bounded to a value $P_{max}$ which complies with the transmit Power Spectral Density regulation:

$$P_{k,n} \leq P_{max}. \quad (2)$$

Given the channel attenuation $a_{k,n}$ experienced by mobile $k$ on subcarrier $n$ (including path loss and multipath fading):

$$P_r(q) \leq a_{k,n}P_{max}. \quad (3)$$

The channel attenuation model on each subcarrier considers free space path loss $a_k$ and multipath Rayleigh fading $a_{k,n}^2$:

$$a_{k,n} = a_k \times a_{k,n}^2. \quad (4)$$

$a_k$ is dependent on the distance between the access point and mobile $k$, $a_{k,n}^2$ represents the multipath fading experienced by mobile $k$ on subcarrier $n$, $a_{k,n}$ is Rayleigh distributed with an expectancy equal to unity.

The maximum number of bits $q_{k,n}$ of mobile $k$ which can be transmitted on a time slot of subcarrier $n$ while keeping below its BER target is:

$$q_{k,n} \leq \left\lfloor \log_2 \left( 1 + \frac{3P_{max} \times T_s \times a_k \times a_{k,n}^2}{2N_0 \left( erfc^{-1} \left( \frac{BER_{target}}{2} \right) \right)^2} \right) \right\rfloor \quad (5)$$

We further assume that the supported QAM modulation orders are limited such as $q$ belongs to the set $S = \{0, 2, 4, \ldots, q_{max}\}$. Hence, the maximum number of bits $B_{k,n}$ that will be transmitted on a time slot of subcarrier $n$ if this resource unit is allocated to the mobile $k$ is:

$$B_{k,n} = \max \{q \in S, q \leq q_{k,n}\}. \quad (6)$$

At each scheduling epoch and for each time slot, MaxSNR based schemes allocate the subcarrier $n$ to the active mobile $k$ which has the greatest $B_{k,n}$ value while the PF scheme consists in allocating the subcarrier $n$ to the mobile $k$ which has the greatest factor $F_{k,n}$ defined as:

$$F_{k,n} = \frac{B_{k,n}}{R_{k,n}}, \quad (7)$$

where $R_{k,n}$ is the time average of the $B_{k,n}$ values.

However, considering rounded $B_{k,n}$ values in the allocation process have a negative discretization side effect on the PF performances. Several mobiles may actually have a same $F_{k,n}$ value with significantly different channel state with respect to their time average. More accuracy is needed in the subcarrier allocation process for prioritizing the mobiles. We think more profitable to allocate the subcarrier $n$ to the mobile $k$ which has the greatest $f_{k,n}$ value such has:

$$f_{k,n} = \frac{b_{k,n}}{r_{k,n}} \quad (8)$$

where:

$$b_{k,n} = \log_2 \left( 1 + \frac{3P_{max} \times T_s \times a_k \times a_{k,n}^2}{2N_0 \left( erfc^{-1} \left( \frac{BER_{target}}{2} \right) \right)^2} \right) \quad (9)$$

and $r_{k,n}$ is the $b_{k,n}$ average over a sliding time window.

PF outperforms MaxSNR providing an equal system capacity and partially improving the fairness. Based on the PF scheme, this paper brings a new scheduler that achieves high fairness while preserving the system throughput maximization.

We introduced a parameter called “Compensation Factor” $CF_k$, that takes into account the current path loss impact on the average possible bit rate of the mobile $k$. It is defined by:

$$CF_k = \frac{b_{ref}}{b_{k,n}}, \quad (10)$$

$b_{ref}$ is a reference number of bits that may be transmitted on a subcarrier considering a reference free space path loss.
\( a_{ref} \) for a reference distance \( d_{ref} \) to the access point and a multipath fading equal to unity:

\[
b_{ref} = \log_2 \left( 1 + \frac{3P_{\text{max}} \times T_s \times a_{ref}}{2N_0 \left( \frac{\text{erfc}^{-1} \left( \frac{\text{BER}_{\text{target}}}{2} \right)}{2} \right)^2} \right). \tag{11}\]

\( b_k \) represents the same quantity but considering a distance \( d_k \) to the access point:

\[
b_k = \log_2 \left( 1 + \frac{3P_{\text{max}} \times T_s \times a_{ref} \times \left( \frac{d_{ref}}{d_k} \right)^2}{2N_0 \left( \frac{\text{erfc}^{-1} \left( \frac{\text{BER}_{\text{target}}}{2} \right)}{2} \right)^2} \right). \tag{12}\]

CPF scheduling consists then in allocating a time slot of subcarrier \( n \) to the mobile \( k \) which has the greatest CPF\(_{k,n}\) value:

\[
\text{CPF}_{k,n} = f_{k,n} \times CF_k = \left( \frac{b_{k,n}}{r_{k,n}} \right) \times CF_k. \tag{13}\]

The distance correction factor \( CF_k \) adequately compensates the lower spectral efficiencies of far mobiles and the resulting \( \text{CPF}_{k,n} \) parameters bring high fairness in the allocation process. Far mobiles get access to the resource more often than close mobiles and inverse proportionally to their spectral efficiency. Thereby, an equal throughput is provided to each mobile. CPF also keeps the PF opportunistic scheduling advantages thanks to the \( f_{k,n} \) parameters which take into account the channel state. Moreover, in contrast with MaxSNR and PF which satisfy much faster the mobiles which are close to the access point, the CPF keeps more mobiles active but with a relatively low traffic backlog. Satisfaction of delay constraints is more uniform and, preserving the multiuser diversity, a better usage of the bandwidth is made. This jointly ensures fairness and system throughput maximization.

The CPF scheduling algorithm is detailed in Fig. 3. The scheduling is performed subcarrier by subcarrier and on a time slot basis for improved granularity. In the allocation process of a given time slot, the priority of a mobile with respect to another is determined by the magnitude of its CPF parameter. In the following, we describe the proposed scheduling algorithm step by step.

- **Step 0:** The scheduler refreshes the current \( b_{k,n} \), \( b_k \) and buffer occupancy \( BO_k \) values and computes the \( f_{k,n} \), \( CF_k \), and \( \text{CPF}_{k,n} \) parameters for each mobile and each subcarrier. Then, \( n \) and \( t \) are initialized to 1.
- **Step 1:** For subcarrier \( n \), the scheduler selects the mobile \( k \) with the greatest \( \text{CPF}_{k,n} \) value.
  - **Sub-step 1-1:** If the virtual buffer occupancy\(^3\) of mobile \( k \) is positive, the scheduler goes to Sub-step 1-2. Else, if all virtual buffers are null or negative, the scheduler goes to Step 2. Otherwise, the scheduler selects the next mobile \( k \) with the greatest \( \text{CPF}_{k,n} \) value and restarts Sub-step 1-1.

\(^3\)We define the virtual buffer occupancy as the current buffer occupancy of mobile \( k \) minus the number of bits already allocated to this mobile.

**IV. Performance Evaluation**

In this section the proposed Compensated Proportional Fair scheduling is compared with the classical Round Robin (RR), the MaxSNR and the PF schemes. Performance evaluation results are obtained using extended OPNET discrete event simulations. We focus on five essential performance criteria: fairness, perceived QoS satisfaction level, mean delay, jitter and offered system capacity.

**A. Simulation setup and QoS criterion**

In the simulations, a frame is composed of 5 time slots and 128 subcarriers. \( (P_{\text{max}} \times T_s)/N_0 \times a_{ref} \) is set to 26.43 dB. Furthermore, in order to study the influence of the distance on the scheduling performances, a first half of mobiles are fixed.
close to the access point at a distance of 1.5 d_{ref}. The second half of mobiles are twice over farther. The $BER_{\text{target}}$ value is taken equal to $10^{-5}$. With these settings, the values of $B_{k,n}$ for the two groups of mobiles are respectively 4 and 2 bits when $\alpha_{k,n}^2$ equals unity.

All mobiles run a videoconference application with successive connections of five minutes duration. The traffic is composed of an MPEG-4 video stream [13] multiplexed with an AMR voice stream [14]. This demanding type of application generates a high volume of data with high sporadicity and requires tight delay constraints which substantially complicates the task of the scheduler. The average bit rate of each source is 80 Kbps. The total number of mobiles sets the traffic load.

A crucial objective for modern multiple access schemes is the full support of multimedia transmission services. Evaluating the QoS offered by a scheduling scheme should not only focus on the classical delay and jitter analysis. Indeed, a meaningful constraint regarding delay is the limitation of the occurrences of large values. In this aim, we define the concept of delay outage by analogy with the concept of outage used in system coverage planning. A mobile transmission is in delay outage when its packets experience a delay greater than a given threshold. The delay experienced by each mobile is tracked all along the lifetime of its connection. At each transmission of a packet of mobile $k$, the ratio of the total number of packets whose delay exceeded the threshold divided by the total number of packets transmitted since the beginning of the connection is computed. The result is called Packet Delay Outage Ratio (PDOR) of mobile $k$ and is denoted $PDOR_k$. Fig. 4 illustrates an example cumulative distribution of the packet delay of a mobile at a given time instant.

The PDOR target is defined as the maximum ratio of packets of mobile $k$ that may be delivered after its delay threshold $T_k$. This characterizes the delay requirements of any mobile in a generic approach. In the following, the PDOR target is set to 5 % and the threshold time $T_k$ is fixed to the value 80 ms considering real time constraints.

B. Simulation Results

Fairness is the most difficult objective to reach. It consists in ensuring the same ratio of packets in delay outage to all mobiles, below the $PDOR_{\text{target}}$. Fig. 5 displays the overall PDOR for various traffic loads. The influence of distance on the scheduling is also studied. An "overall" PDOR value is computed for all packets transmitted by the mobiles situated close to the access point in Fig. 5(a), and for the mobiles situated twice over farther in Fig. 5(b).

Classical RR yields bad results. Indeed, since multiuser diversity is not exploited, the overall spectral efficiency is small and system throughput is low. Consequently, the delay targets are exceeded as soon as the traffic load increases. Based on opportunistic scheduling, MaxSNR, PF and CPF provide better system performances. However, with MaxSNR and PF, close mobiles easily respect their delay requirement but the farther experience much higher delays and go beyond their PDOR target when the traffic load increases. This shows their difficulty to ensure fairness with respect to different mobile positions. The problem is solved with CPF which provides comparable QoS levels to all mobiles whatever their respective location and allows to reach higher traffic loads with acceptable PDOR (below the PDOR target). If we focus on Fig. 5(c), we notice that, besides ensuring high fairness, CPF provides a better overall QoS level as well.

We then examined the mobile satisfaction with respect to delay. We consider a mobile is satisfied when, at the end of its connection, its delay constraint is met (i.e. $PDOR \leq PDOR_{\text{target}}$). Fig. 6(a) shows the percentage of mobile connections that do not satisfy the PDOR target. Clearly the CPF brings the largest satisfaction even under high traffic load. Fig. 7 further details the analysis and shows the CDF of end connection PDOR values for a traffic load of 960 Kbps, using respectively the MaxSNR, the PF and the CPF schemes.

Highly unfair, MaxSNR fully satisfies the required QoS of close mobiles at the expense of the satisfaction of far mobiles. Indeed, only 54.5 percents of these latter experience a final PDOR inferior to the PDOR target of 5 % (cf. Fig. 7(a)). Unnecessary priorities are given to close mobiles who easily respect their QoS constraints while more attention should be given to the farther. These inadequate priority management dramatically increases the global mobile dissatisfaction which reaches 23 % as shown on Fig. 6(a) and Fig. 7(a).

PF brings more fairness and allocates more priority to far mobiles. Compared to MaxSNR, PF offers a QoS support improvement with only 12.8 % of dissatisfied mobiles (cf. Fig. 6(a) and Fig. 7(b)). This result indicates that some flows can be slightly delayed to the benefit of others without significantly affecting their QoS.

The CPF was built on this idea. The easy satisfaction of close mobiles (with better spectral efficiency) offers a degree of freedom which ideally should be exploited in order to help the farther ones. CPF dynamically adapts the priorities function of the mobile location. This results in allocating to each mobile the accurate share of bandwidth required for the...
satisfaction of its QoS constraints, whatever its position. With CPF, only 5.3 percents of the mobiles are dissatisfied, i.e. half less than with PF and four times less than with MaxSNR (cf. Fig. 6(a) and Fig. 7(c)). Additionally, compared to Fig. 7(a) and Fig. 7(b), Fig. 7(c) exhibits superimposed curves which proves the CPF high fairness.

The good results of CPF are corroborated in Fig. 6(b) and 6(c) where the overall values of the mean packet delay and jitter obtained with CPF are minimized.

We finally looked at the cost of this high fairness and mobile satisfaction in terms of system capacity. Bandwidth waste was first analyzed. We consider a resource unit is wasted when the scheduler cannot allocate this resource unit to any mobile. This is the case when all active mobiles experience a strong multipath fading on the considered subcarrier and all modulation orders fail at ensuring the BER target. Fig. 8 shows the percentage of resource units which are wasted in the total number of resource units considered for allocation. These results give a good idea of the share of bandwidth wasted by each scheduler. RR which does not take advantage of multiuser diversity suffers of a constant and high bandwidth waste whatever the traffic load. In contrast, the other schemes use multiuser diversity and avoid waste. The greater the number of active mobiles, the greater the possible choice among mobiles with limited multipath fading and the lower the bandwidth waste.

The bandwidth usage ratio, defined as the mean number
of allocated resource units divided by the total number of resource units, is shown in Fig. 9. It appears that no system throughput reduction has been done with CPF. As expected with the non opportunistic RR, the bandwidth usage ratio is proportional to the traffic load. In contrast, with the opportunistic schedulers (MaxSNR, PF, CPF) an inflection of the bandwidth usage ratio curves appears when the traffic load increases. Actually, in these simulations, the traffic load growth corresponds to an increase of the number of mobiles. The curve inflection shows that the opportunistic schedulers take advantage of this supplementary multiplex user diversity for optimizing the bandwidth usage efficiency.

Carefully observing the above performance results, the CPF scheduling has slightly better performances than the two other opportunistic schedulers. Indeed, CPF makes less waste (Fig. 8) and, at the highest traffic load of 1280 Kbps, CPF keeps 1.5 % more bandwidth available than the MaxSNR though acknowledged as the reference scheduler in terms of system capacity maximization (Fig. 9). While ensuring fairness, CPF makes a well-balanced resource allocation which keeps a higher number of mobiles active across time and with continuously low traffic backlogs. Preserving this multiuser diversity allows to continuously take a maximal benefit of opportunistic scheduling and thus maximize the bandwidth usage efficiency better than state of the art scheduling schemes like MaxSNR and PF.

V. Conclusion

In the litterature, several scheduling schemes have been proposed for maximizing the system throughput. However, guaranteeing a high fairness appeared as unfeasible without sacrificing system capacity. In this paper, we proposed an improvement of the PF scheduling scheme yet acknowledged as the most promising so far. This scheme, called “Compensated Proportional Fair (CPF)”, allows to avoid the tradeoff between fairness and system capacity. It has a low complexity and is easily implementable on all OFDM based networks like 802.11 a/g and 802.16 networks. CPF sparingly delays the flows of close mobiles with good spectral efficiency in order to favor the flows of the farther mobiles which need more attention for fulfilling their delay constraints. Performance results show that CPF provides both high fairness and system throughput maximization making a better usage of multiuser diversity.

References