RESEARCH   **Open Access**

Traffic models impact on OFDMA scheduling design

Pablo Ameigeiras[1], Yuanye Wang[2], Jorge Navarro-Ortiz[1], Preben E Mogensen[2,3] and Juan M Lopez-Soler[1*]

Abstract

This article studies the impact on the design of scheduling algorithms for *Orthogonal Frequency Division Multiple Access* (OFDMA) systems of two traffic models described in the evaluation methodology proposals from standardization bodies: the full buffer and the finite buffer traffic models. The analysis concentrates on utility-based scheduling with an α-fair utility function for *Non-Real Time* (NRT) services. The results show that a gradient scheduling algorithm is able to maximize the aggregate utility over all the users when the less realistic full buffer model is adopted; but not when the finite buffer model is applied. The results also show that with the full buffer model a gradient scheduler exhibits a trade-off between average user throughput and the user throughput at 5% outage, but it does not when the more realistic finite buffer is used. Therefore, it is concluded that designs of scheduling algorithms for NRT services for OFDMA systems carried out under the full buffer model assumption may fail to provide the desired performance benefits in realistic scenarios. Based on the results presented, a recommendation on scheduling design is provided.

Keywords: finite buffer traffic model, full buffer traffic model, OFDMA, packet scheduling, α-fair function, fairness.

1. Introduction

Orthogonal Frequency Division Multiple Access (OFDMA) has become an essential technology in broadband wireless communications. Systems such as the 3GPP *Long-Term Evolution* (LTE) [1], IEEE 802.16d (i.e. fixed WiMAX) [2], IEEE 802.16e (i.e. mobile WiMAX) [3], LTE-Advanced [4] and IEEE 802.16m [5] have adopted OFDMA as their multiple access technique for either downlink or both uplink and downlink air interfaces.

In the design of the mentioned wireless access systems, the evaluation methods and models play a key role. There exist publications in the literature that have already indicated that the performance results of algorithms [6] or enhancing features [7] in OFDMA systems significantly depends on the considered traffic model. In this respect, the *Next Generation Mobile Networks* (NGMN) Alliance and the *Institute of Electrical and Electronic Engineers* (IEEE) have proposed evaluation methodologies for next generation wireless access

networks in [8,9], respectively. In addition, evaluation models from the *3rd Generation Partnership Project* (3GPP) can be found, for example, in [10]. In these documents, two types of simplified traffic models can be identified for *Non-Real Time* (NRT) services:

(i) *Full buffer model* [8-10]: it is a simplified version of the traffic received/transmitted by a user in a data session. It is characterized by two facts: the number of users in the cell is constant and the buffers of the users' data flows always have unlimited amount of data to transmit. The full buffer model has widely been adopted in OFDMA in simulation-based [11-14] and theoretical [15-17] investigations due to its simplicity.

(ii) *Finite buffer model* [8,10] (in [10] the so-called *FTP traffic model 1*): it is a simplified version of the interactive or background traffic type. It includes both a user arrival (birth) and a user departure (death) process. With this traffic model, a user is assigned a finite payload to transmit or receive when it arrives, and it leaves the system after the payload transmission or reception is completed. The user arrival process of this model captures the fact that the users in the network are not simultaneously active at the same time, but they rather become active when they start a data session that

* Correspondence: juanma@ugr.es
[1]Department of Signal Theory, Telematics and Communications, University of Granada, Granada, Spain
Full list of author information is available at the end of the article

requires the download/upload of data [18-21]. This model has been less extensively adopted [22,23].

The evaluation methodology documents [8,9] have also proposed more complex models based on statistical distributions where users are assumed to execute applications (e.g. FTP, Web browsing, etc.). However, these models do not include a user arrival or departure process and have been much less extensively adopted due to their increased complexity.

Many investigations about scheduling in OFDMA systems have been carried out only under the full buffer model [14,16,17,24,25] and therefore their results and conclusions are influenced by the full queue effect. This article studies the influence of the traffic model on the performance of packet scheduling algorithms in OFDMA systems. In particular, we concentrate on utility-based scheduling [17,26] for two reasons: first, it has been reported that it provides an effective trade-off among spectral efficiency and fairness in wireless networks [14,27]; and second, proportional fairness scheduling [28], which is a particular case of utility-based scheduling [26], has widely been adopted in OFDMA systems [24,29], and it is the basis for some QoS scheduling algorithms [16,30].

The evaluation of the traffic model impact will be carried out by means of quasi-dynamic system-level simulations. For the evaluation, two OFDMA network models are considered: (i) initially, a simple OFDMA network model. This simple network model assumes an ideal achievable transmission rate per hertz through the Shannon's law, and that the base station has ideal knowledge of the user's *Signal-to-Interference Ratio* (SIR). (ii) Later, a realistic 3G LTE network model will be considered. It will include relevant features in an LTE system such as the *Modulation and Coding Schemes* (MCS), Hybrid-ARQ and the *Channel Quality Information* (CQI) feedback. The evaluation with these two network models will allow us to determine if the effect of the traffic models on the performance of a realistic OFDMA network model, such as 3G LTE, significantly differs from the case of a basic OFDMA network model.

The obtained results will show that under the OFDMA network model:

A gradient scheduling algorithm is able to maximize the aggregate utility over all the users (for a given utility function) when the less realistic full buffer model is adopted; but not when the finite buffer model is applied.

With the full buffer model, a gradient scheduling algorithm exhibits a trade-off between average user throughput and the user throughput at 5% outage. However, when the more realistic finite buffer is used, this trade-off does not exist.

The relative comparison of key performance indicators (KPIs), such as the average user throughput or the user throughput at 5% outage, between versions of a gradient scheduler with different fairness degrees is insensitive to the network load under the full buffer traffic model, but sensitive under the finite buffer traffic model.

Moreover, with the realistic 3G LTE network model, the same trends are found as with the simple OFDMA network model.

Based on these results, it is concluded that designing scheduling algorithms for NRT services for OFDMA systems under the full buffer model assumption may not yield the desired performance benefits in realistic scenarios where the finite buffer model is more appropriate due to its birth and death process.

As a consequence of these findings, we propose a scheduling algorithm based on the gradient scheduler combined with a simple utility function. This proposal is able to achieve a similar user throughput at 5% outage compared to the Proportional Fair scheduler but an average user throughput gain that reaches up to 60-70% for heavy offered load conditions.

The rest of this article is organized as follows. Section 2 describes the simple OFDMA system model and Section 3 presents the adopted scheduling algorithm based on Utility Theory. Section 4 presents the full buffer and the finite buffer traffic models as well as KPIs. Section 5 evaluates the influence of the full and finite buffer traffic models on an OFDMA network. Additionally, Section 6 includes a similar analysis in a realistic 3G LTE network. Finally, Section 7 draws the main conclusions.

2. Simple OFDMA system model

Let us consider the downlink transmission of an OFDMA cell where a base station transmits data to a set of users. Let $M = \{1,...,M\}$ represent both the set of users and its cardinality at an arbitrary epoch in time. The base station carries out the transmission towards the M users through an N-point IFFT in which time is slotted in *Transmission Time Intervals* (TTI) of fixed duration T at a carrier frequency f_c. A total bandwidth B is available for transmission composed of N_u useful subcarriers with a subcarrier spacing S_s. The transmission considers a *Single Input Multiple Output* configuration with two receiving antennas. Assuming perfect sample and synchronization and sufficient cyclic prefix length, the discrete-time baseband signal received by user k on subcarrier s, TTI n, and antenna i can be expressed as:

$$y_k^i[n, s] = H_k^i[n, s] \cdot \sqrt{p[n, s]} \cdot x_k[n, s] + n_k^i[n, s] + z_k^i[n, s] \quad (1)$$

where $H_k^i[n, s]$ denotes the N-point DFT of the complex baseband representation of the channel impulse response:

$$H_k^i[n,s] = \sum_{l=1}^{L} a_k^i[n,l] \cdot e^{-j\frac{2\pi ls}{N}} \quad (2)$$

where $a_k^i[n,l]$ being the lth time domain channel tap experienced by the user k at antenna i. The channel taps include the small scale fading, shadow fading and path loss, and are assumed to be stationary narrowband, complex Gaussian processes, which are independent for different paths, users and antennas at the receiver. The time-domain channel taps are assumed to be constant for the duration of the TTI, but are updated every TTI with a maximum Doppler frequency f_d. $p[n,s]$ is the base station transmission power on subcarrier s, and it is assumed to be constant across subcarriers, i.e. $p[n,s] = p[n] \, \forall s$. The transmitted data symbols $x_k[n,s]$ are independent random variables with zero mean and power $E\{|x_k[n,s]|^2\} = 1$. The signal is corrupted with zero-mean AWGN noise samples $n_k^i[n,s]$ with variance:

$$E\left\{\left|n_k^1[n,s]\right|^2\right\} = E\left\{\left|n_k^2[n,s]\right|^2\right\} = \frac{N_0 \cdot B}{N} = \sigma_n^2 \quad (3)$$

where N_0 represents the one-sided power spectral density of the noise. A hexagonal grid with J surrounding cells is considered for the generation of time continuous interference. For simplicity purposes, the interference samples $z_k^i[n,s]$ are modeled as zero-mean AWGN samples whose variance:

$$E\left\{\left|z_k^1[n,s]\right|^2\right\} = E\left\{\left|z_k^2[n,s]\right|^2\right\} = \sum_{j=1}^{J} I_{j,k}[n,s] \quad (4)$$

is computed as the sum of the interference power $I_{j,k}[n,s]$ received by user k from every jth interfering cell based on the deterministic path loss, shadow fading and interfering base station transmission power.

The resulting *Signal-to-Noise plus Interference Ratio* $\text{SINR}_k^i[n,s]$ of every user k in every subcarrier s and antenna i can be computed as:

$$\text{SINR}_k^i[n,s] = \frac{p[n,s] \cdot \left|H_k^i[n,s]\right|^2}{\sigma_n^2 + \sum_{j=1}^{J} I_{j,k}} \quad (5)$$

Assuming ideal *Maximal Ratio Combining* (MRC) at the receiver, the SIR at the detector of every user k in every subcarrier s can be computed as:

$$\text{SINR}_k[n,s] = \text{SINR}_k^1[n,s] + \text{SINR}_k^2[n,s] \quad (6)$$

The achievable transmission rate $R_k[n,s]$ of user k in subcarrier s and TTI n is computed with the Shannon law, but including an efficiency factor due to the *Cyclic Prefix* (CP) of the OFDM signal:

$$R_k[n,s] = \rho \cdot W \cdot \log_2\left(1 + \text{SINR}_k[n,s]\right) \quad (7)$$

where $R_k[n,s]$ denotes the achievable transmission rate, W represents the subcarrier bandwidth and ρ is the CP efficiency factor.

The base station is assumed to have instantaneous and perfect knowledge of the achievable transmission rate $R_k[n,s]$ of all the users on every subcarrier and at every TTI. The scheduling algorithm uses this information to assign bandwidth resources to the users. The granularity of the bandwidth assignment equals 1 subcarrier:

$$\psi_k[n,s] = \begin{cases} 1, & \text{subcarrier } s \text{ is assigned to user } k \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

There is no user multiplexing per subcarrier, and hence at maximum one user is assigned on each subcarrier. In equation $\sum_{\forall k \in M} \psi_k[n,s] \in \{0,1\}$, note that there can be unassigned subcarriers only when the finite buffer model is used. With the full buffer model, there is always data to transmit, and $\sum_{\forall k \in M} \psi_k[n,s] = 1$.

Whenever a user k is assigned subcarrier s, he is assumed to correctly receive the following amount of data:

$$d_k[n,s] = R_k[n,s] \cdot T \quad (9)$$

where T is the TTI duration.

At the *Medium Access Control* layer of the base station, the queue $q_k[n]$ corresponding to the flow of the user k is decremented according to the transmitted bits:

$$q_k[n] = q_k[n-1] - \sum_{s=1}^{N_u} d_k[n,s] \cdot \psi_k[n,s] \quad (10)$$

3. Packet scheduling based on utility theory

Scheduling algorithms based on Utility Theory have widely been applied in OFDMA networks due to their promising performance and implementation simplicity. These algorithms incorporate the level of customer satisfaction in the resource allocation through the use of utility functions that may be obtained from surveys of the users' subjective experienced quality (like the one in [31]), or may be designed based on the traffic habits or desired fairness in the network [32]. The Utility Theory proposes the optimization objective of maximizing the aggregate utility over all users in a cell for a given utility function $U(\cdot)$ [26]. For NRT traffic, the considered optimization objective is the maximization of the aggregate utility with respect to the user throughput r_k:

$$\max \sum_{k \in M} U(r_k) \qquad (11)$$

where $U(\cdot)$ is a utility function that quantifies the degree to which a network satisfies the user demands of NRT applications [32] based on the user throughput r_k.

Song and Li [17] developed a cross-layer radio resource assignment framework for NRT traffic in OFDMA networks that aims at solving objective (11). In their framework, they studied dynamic subcarrier assignment and adaptive power allocation schemes. Concentrating on the dynamic subcarrier assignment, Song and Li [17] demonstrated that, assuming a constant power allocation per subcarrier, the maximization of the aggregate utility, i.e. objective (11), can be achieved by a gradient scheduler algorithm that assigns subcarrier s in TTI n to the user m that fulfills:

$$m[n, s] = \arg \max_{k \in M} \left[U'\left(\overline{r_k[n-1]} \right) \cdot R_k[n, s] \right] \qquad (12)$$

where

$$U'\left(\overline{r_k[n-1]} \right) = \frac{\partial U(r)}{\partial r} \Bigg|_{r = \overline{r_k[n-1]}} \qquad (13)$$

In (12), the user throughput has been approximated by the average data rate during a certain period $\overline{r_k[n]}$, which is calculated using an exponentially weighted low-pass time window:

$$\overline{r_k[n]} = (1 - \rho_w) \cdot \overline{r_k[n-1]} + \rho_w \cdot r_k[n] \qquad (14)$$

where $\rho_w = T/T_w$, T_w is the length of the averaging window, and $r_k[n]$ is the actual transmitted data rate for user k at time n across its assigned subcarriers.

Let us consider the α-fair function as the utility function to be used in (11). The α-fair function is a family of functions defined in [33] as:

$$U_\alpha(r) = \begin{cases} \dfrac{r^{1-\alpha}}{1-\alpha} & \alpha \geq 0, \, \alpha \neq 1 \\ \log(r) & \alpha = 1 \end{cases} \qquad (15)$$

where $U_\alpha(\cdot)$ is the α-fair function.

As described in [14,27,34,35], the application of the α-fair function as a utility function in the objective (11) leads to a resource allocation where the parameter α controls the degree of fairness of the allocation. When $\alpha = 0$, the network throughput is maximized at the expenses of the throughput starvation of some users, when $\alpha = 1$, the throughput allocation is proportionally fair, and when $\alpha \to \infty$ the allocation corresponds to max-min fairness.

Then, assuming that the α-fair function is selected as utility function, the gradient scheduling algorithm that maximizes the aggregate utility over all users should compute the priority of user k in subcarrier s and TTI n as:

$$\begin{aligned} P_k[n, s] &= \frac{\partial U_\alpha(r)}{\partial r} \Bigg|_{r = \overline{r_k[n-1]}} \cdot R_k[n, s] \\ &= r^{-\alpha} \Big|_{r = \overline{r_k[n-1]}} \cdot R_k[n, s] \\ &= \frac{R_k[n, s]}{\left[\overline{r_k[n-1]} \right]^\alpha} \end{aligned} \qquad (16)$$

for $\alpha \geq 0$. This scheduling discipline will be hereafter referred to as α-fair gradient scheduler.

It is relevant to mention that the optimal solution of objective (11) changes over time because both the channel capacity and the set of active users M vary with time [36]. In fact, the scheduling decisions determine the user departure process, which ultimately influences the set of active users M. Therefore, it is questionable if the optimal solution to the objective (11) over a long observation period is to continuously target at the "present" optimal solution by moving in the direction of the gradient as in (12).

4. Description of the full buffer traffic model, finite buffer traffic model, and KPIs
4.1. Full buffer traffic model
The so-called *full buffer traffic model* is widely used in the performance evaluation of wireless networks [11-13]. It is characterized by the fact that the buffers of the users' data flows always have unlimited amount of data to transmit [9] and, therefore, transmission of their data payloads never finishes. This implies that $q_k[n] = \infty \; \forall k$ and $\forall n$ in Equation (10). In addition, there is no user arrival process in the full buffer model and, therefore, the number of users does not vary.

In simulation-based evaluations, typically, during a simulation run the locations of the users are kept invariable and therefore multiple independent runs (*snapshots*) are required with the objective of sampling the geographical distribution of the users in the footprint area. At the end of the process, the statistics are collected from all the simulated snapshots. In this study, it will be assumed that the number of users in the cell for each snapshot follows a Poisson distribution.

In theoretical investigations, this model allows a simplified analysis.

4.2. Finite buffer traffic model
The so-called *finite buffer traffic model* is characterized by the fact that the users have a finite amount of data to transmit or receive. Once the data transmission is finished, the session is also finished and the user is

removed from the system. For simplicity purposes, the session is typically composed of a single data payload of a given size [36], and for the same reason, the effects of higher layer protocols such as TCP, HTTP, are not considered. The model includes a process of user arrivals of a given rate λ that provides the instants when the users arrive in the system and start their data transmission or reception. This process models the fact that all users in the OFDMA network are not simultaneously active, but they rather become active when they demand the download/upload of data [18-21]. This study models the user arrivals with a Poisson process (i.e. exponentially distributed inter-arrival times). Note that the sum of a large number of independent stationary renewal processes tends to a Poisson process [37].

With the finite buffer traffic model, a user stays in the system from the instant of the arrival until the data transmission is finished. This period depends on the throughput achieved by the user, and therefore the instantaneous number of users in the system depends on how the data rates are assigned by the base station. With this model, as the number of users in the cell instantaneously varies with the time, there exist temporary periods when the number of users reach crests caused by peaks of user arrivals or accumulation of very poor channel quality users.

In simulation-based evaluations, with the finite buffer traffic model typically a single simulation run is executed [38], but it should be long enough to collect sufficient statistics for accurate sampling of relevant performance metrics. Sampling the load peaks requires a large amount of statistics and therefore increases the complexity compared to the full buffer model.

We consider that the results obtained under the finite buffer model are more generalizable than those obtained with the full buffer because the latter disregards the fact that in a realistic OFDMA network all users do not simultaneously request the download/upload of an infinite amount of data.

4.3. Overview of KPIs

The selection of the KPIs is a decisive issue because the final scheduling design will depend on them. This study makes use of KPIs proposed in [8-10] that are widely applied in performance evaluations of OFDMA systems because they provide information about the average spectral efficiency and throughput outage levels. For convenience, the KPIs will be defined as relative gains over the Proportional Fair case:

User throughput r_k. Let denote $n_{k,a}$ and $n_{k,d}$ the arrival and departure TTIs of user k in the OFDMA system. The user throughput r_k is computed as the ratio of the number of data bits successfully received by user k divided by the elapsed time between arrival and

departure:

$$r_k = \frac{1}{T \cdot (n_{k,d} - n_{k,a} + 1)} \cdot \sum_{n=n_{k,a}}^{n_{k,d}} \sum_{s=1}^{N_u} d_k[n,s] \cdot \psi_k[n,s] \quad (17)$$

Average user throughput gain over Proportional Fair G_Ω. Let $K = \{1,..., K\}$ represent both the cardinality and the set of all users that arrive in the system over the observation period. Let compute the average user throughput provided by the scheduling algorithm in (16) with parameter α over the set of users K as $\Omega_\alpha = \frac{1}{K} \sum_{k=1}^{K} r_k$. Then, we define the average user throughput gain of over Proportional Fair as:

$$G_\Omega = {\Omega_\alpha}/{\Omega_1} \quad (18)$$

User throughput at 5% outage gain over Proportional Fair G_O. Let denote O_α the fifth percentile of the *Cumulative Distribution Function* (CDF) of the user throughput provided by the scheduling algorithm in (16) with parameter α. O_α corresponds to the 5% worst user throughput. Then, we define the user throughput at 5% outage gain over Proportional Fair as

$$G_O = {O_\alpha}/{O_1} \quad (19)$$

In addition, a utility metric, that assesses the degree of user satisfaction, will be considered for the evaluation:

Average utility gain over Proportional Fair G_U: Let u_k denote the utility of user k with a throughput r_k, $u_k = U(r_k)$ for a given utility function $U(\cdot)$. Let compute the average utility provided by the scheduling algorithm in (16) with parameter α over the set of users K as $U_\alpha = \frac{1}{K} \sum_{k=1}^{K} u_k$. Then, we define the average utility gain over Proportional Fair as:

$$G_U = {U_\alpha}/{U_1} \quad (20)$$

5. Performance evaluation under a simple OFDMA network model

5.1. Experimental set-up

The OFDMA system model presented in Section 2 has precisely been reproduced in a quasi-dynamic network simulator. Users are uniformly distributed in area, they do not to vary their geographical location with time, and therefore, their deterministic path loss and shadow fading remains invariable during the lifetime of the user. The small scale fading of the channel taps is modeled according to the ITU *Typical Urban* (TU) power delay

profile [39] assuming a sampling frequency f_s = 15.36 MHz.

For the full buffer model, a total of 1,000 snapshots are executed to have an accurate sampling of the geographical distribution of the users in the cell as well as the distribution of the number of users (i.e., Poisson distribution). Different average cell loads are achieved by varying the average of the Poisson distribution. Every snapshot is set to last 10 s, which is found to be sufficient for the relevant statistics to converge.

For the finite buffer model, in order to be able to analyze the levels of average offered load close to the maximum cell capacity, 30 independent runs (with different seeds) are executed, with 1500 s of duration for each run. As in [29], here it is also assumed that the size of the user's payload is equal to 2 Mbits. The remaining parameters are included in Table 1.

5.2. Performance results assuming a fixed network load

This section presents the performance of the α-fair gradient scheduling algorithm under the finite and full buffer traffic models for different α values (see (16)) assuming a fixed offered load. For the comparison, the Poisson distribution of the full buffer traffic model is set to have an average number of users in the cell equal to 15. For the finite buffer, the input of the model is the average user arrival rate which decides the offered cell load of the Poisson arrival process. Here, the offered cell load is selected to be equal to 23 Mbps because it yields a load level in the system (in terms of average number

Table 1 OFDMA network model

Parameter	Setting
Carrier frequency	2 GHz
Transmission bandwidth	10 MHz
OFDMA subcarrier spacing	15 kHz
Number of useful subcarriers	600
Number of points of FFT	1024
Cyclic prefix length	72 samples
TTI duration (simulation step)	1 ms
Sampling frequency	15.36 MHz
Number of OFDM symbols per TTI	14
Inter-site distance	2 km
Path loss exponent	3.5
Std of shadow fading	8 dB
Channel model	ITU typical urban 20 paths
Mobile speed	3 km/h
Base station transmission power	40 W
Number of interfering cells	12
Antenna rx scheme	2-rx MRC
Bandwidth granularity of scheduling assignments	1 subcarrier

of users) in the same order of magnitude as the one used in the case of the full buffer model. This criterion will also be used in the following sections to decide the offered cell load for the finite buffer model.

Figure 1 depicts the CDF of the user throughput for two selected α values, α = 1 and α = 0.6, under the full and finite buffer traffic models. Under the full buffer model, the CDFs of the user throughput for both α values cross at approximately the 60th percentile of the CDF. This trend is expected because by decreasing the value of α from 1 to 0.6, the scheduling algorithm becomes more unfair, favoring users under good channel conditions at the expenses of the remaining users. However, under the finite buffer model, the CDFs of the user throughput for both α values cross at approximately 10th percentile and the tails of both CDFs approximate to each other. The main reason behind this difference in the trend is that, under the finite buffer traffic model, increasing the priority of users under good channel conditions does not always degrade the throughput of users with poor channel conditions because the former ones finish their data transmissions and departure from the system thereby releasing the bandwidth resources.

Figure 2a depicts the average user throughput gain over Proportional Fair G_Ω. The results show that for $\alpha < 1$, the finite buffer traffic model presents a much higher average user throughput gain than the full buffer model.

Figure 2b depicts the user throughput at 5% outage gain over Proportional Fair G_O. For the full buffer traffic model, the user throughput at 5% outage increases for higher values of α, thereby, as expected, improving the degree of fairness provided by the α-fair gradient scheduling algorithm (as already shown in [14]). On the other hand, when the finite buffer model is used, the user throughput at 5% outage exhibits a maximum at $\alpha \approx [0.75\text{-}1]$, and α values outside this range cause a performance loss. The reason is that, with the finite buffer model, giving more resources to users under favorable channel quality conditions ($\alpha < 1$) reduces their alive time, and therefore users with poor channel quality can get sufficient resources after the departure of the favorable ones, whereas, in the case of full buffer model, users are always competing for resources.

Then, the results show that under the full buffer traffic model there exists a trade-off between the average user throughput and the user throughput at 5% outage, but under the finite buffer model this trade-off does not exist because the user throughput at 5% outage presents an optimum performance in the range $\alpha \approx [0.75\text{-}1]$.

Figure 3a depicts the average utility gain over Proportional Fair G_U for different α values assuming that the utility function is the natural logarithm:
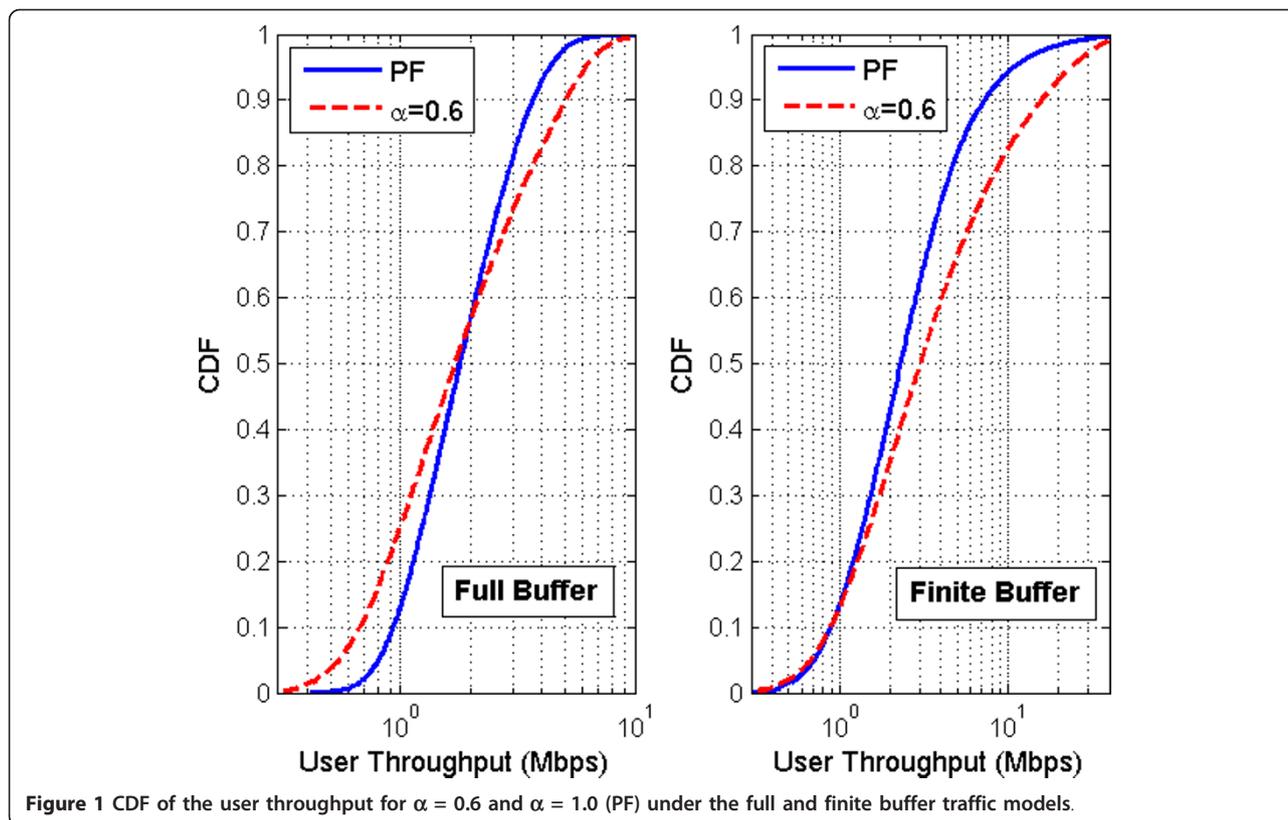
Figure 1 CDF of the user throughput for α = 0.6 and α = 1.0 (PF) under the full and finite buffer traffic models.

$$U(r_k) = \log(r_k) \tag{21}$$

where r_k denotes the throughput of user k. As expected, under the full buffer model the gradient scheduling algorithm with $\alpha = 1$ maximizes the average utility when the utility function is (21). As the average utility is maximized, the aggregate utility $\sum_{k \in M} U(r_k)$ is also maximized. However, when the finite buffer traffic model is used, the maximization of the average utility does not hold for $\alpha = 1$, but it is rather increased as $\alpha \to 0$. Figure 3b depicts the average utility when the utility function is the α-fair function with $\alpha = 0.5$ (see (22)) reaching a similar conclusion.

$$U(r_k) = 2 \cdot \sqrt{r_k} \tag{22}$$

Consequently, the results show that the α-fair gradient scheduling algorithm maximizes the average utility as long as the full buffer traffic model is used. But, when the finite buffer traffic model is used, the gradient scheduling algorithm does not keep this property.

5.3. Performance results under different network load conditions

This section analyzes the sensitivity of the performance of the α-fair gradient scheduler to the network load under the finite and full buffer traffic models. For this purpose, two versions of the α-fair gradient scheduling algorithm will be evaluated under various cell loads: the first is $\alpha = 1$, which corresponds to the well-known Proportional Fair scheduling; the second is $\alpha = 0.6$, which provides interesting results from the viewpoint of the combined performance of the average and 5% outage user throughput. Both scheduling algorithms are evaluated under the full buffer traffic model with the following average number of users in the cell $N = [8, 11, 14, 17, 20]$. For the finite buffer model, the following loads are considered $L = [20, 21, 22, 23, 24, 25]$, where L represents the offered cell load measured in Mbps.

Figure 4 represents the evolution of the average user throughput gain over Proportional Fair G_Ω and user throughput at 5% outage gain over Proportional Fair G_O for both traffic models as a function of the cell load. It is interesting to observe that G_Ω and G_O achieved with the α-fair gradient scheduling algorithm with $\alpha = 0.6$ remains approximately insensitive to the load under the full buffer traffic model. But, this behavior does not hold for the finite buffer traffic model. For example, with offered cell loads of 20 and 25 Mbps, the relative gains of the average user throughput equal 1.49 and 1.81, respectively. With the same offered loads, the relative gains of the 5% outage throughput equal 0.94 and 1.0, respectively.
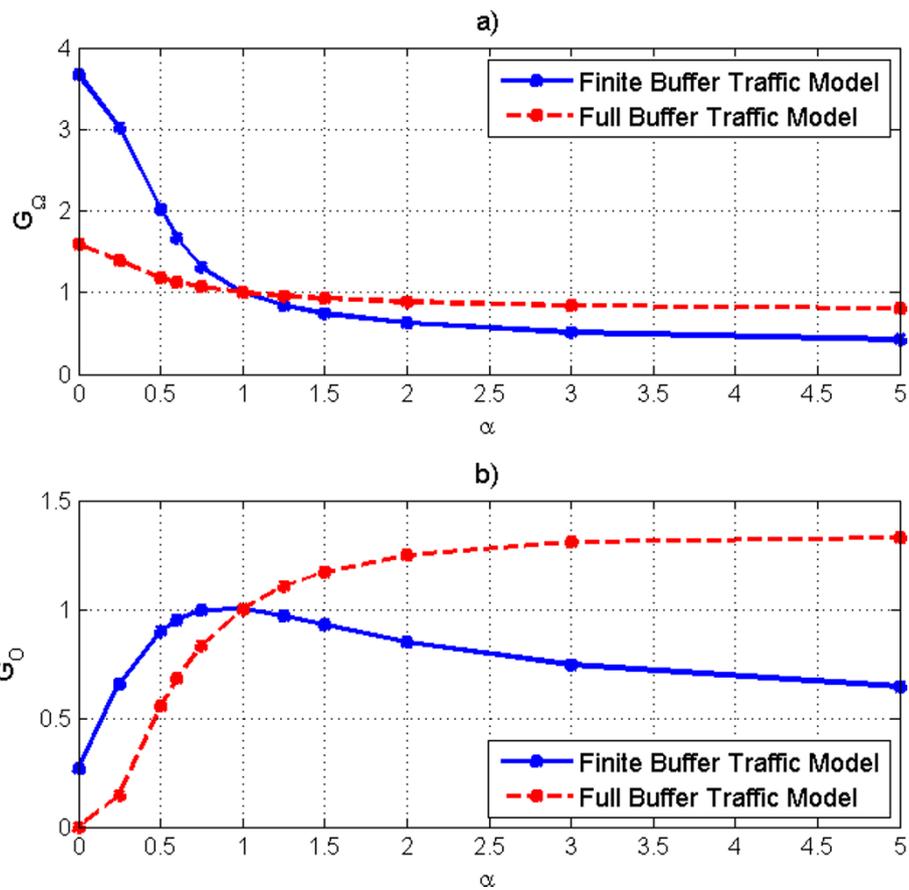
Figure 2 Finite and full buffer traffic models comparison. **(a)** Average user throughput gain over Proportional Fair G_Ω. **(b)** User throughput at 5% outage gain over Proportional Fair G_O for different α values.

Hence, the results show that the relative comparison of KPIs, such as the average user throughput or the user throughput at 5% outage, between two different α values of the α-fair gradient scheduling algorithm, is insensitive to the network load under the full buffer traffic model. However, the same relative comparison is indeed load sensitive under the finite buffer traffic model.

6. Performance evaluation under a realistic OFDMA network model: 3G LTE

The previous section has shown that the α-fair gradient scheduling algorithm behaves significantly different with full or finite buffer model under a simple OFDMA system model. This section aims at providing the evaluation with realistic system-level assumptions. For this purpose, the downlink of an LTE system is taken as a study case. The performance is also investigated here with various load conditions.

The OFDMA simulator described in Section 5.1 has been modified to follow the 3G LTE specifications defined in [1]. The propagation, user location and mobility, and traffic models are kept as in the OFDMA

network model. The transmission bandwidth and the base station power are also maintained. However, the link to system mapping is based on the exponential effective SIR mapping model [40], and a finite set of MCS is used instead of Shannon's law. A HARQ protocol with 6-Stop and Wait (SAW) channels is included to retransmit the erroneously decoded radio blocks. The scheduler assigns resource units of 180 kHz and 1 ms. An Outer Loop Link Adaptation algorithm is also included to adjust the Block Erasure Rate (BLER) to a target value. The modeling of the channel quality indicator and its feedback (CQI) is extracted from [41]. The number of runs and the duration of each run are also maintained as in Section 5. Other 3G LTE specific parameters and assumptions are summarized in Table 2.

6.1. Performance results assuming a fixed network load

As in previous section, we choose again $\alpha = 0.6$ for the study with realistic assumptions and Proportional Fair ($\alpha = 1$) as reference. For the full buffer model, the average number of users is set $N = 20$, which is selected to be higher than in Section 5.1 ($N = 15$) due to the better
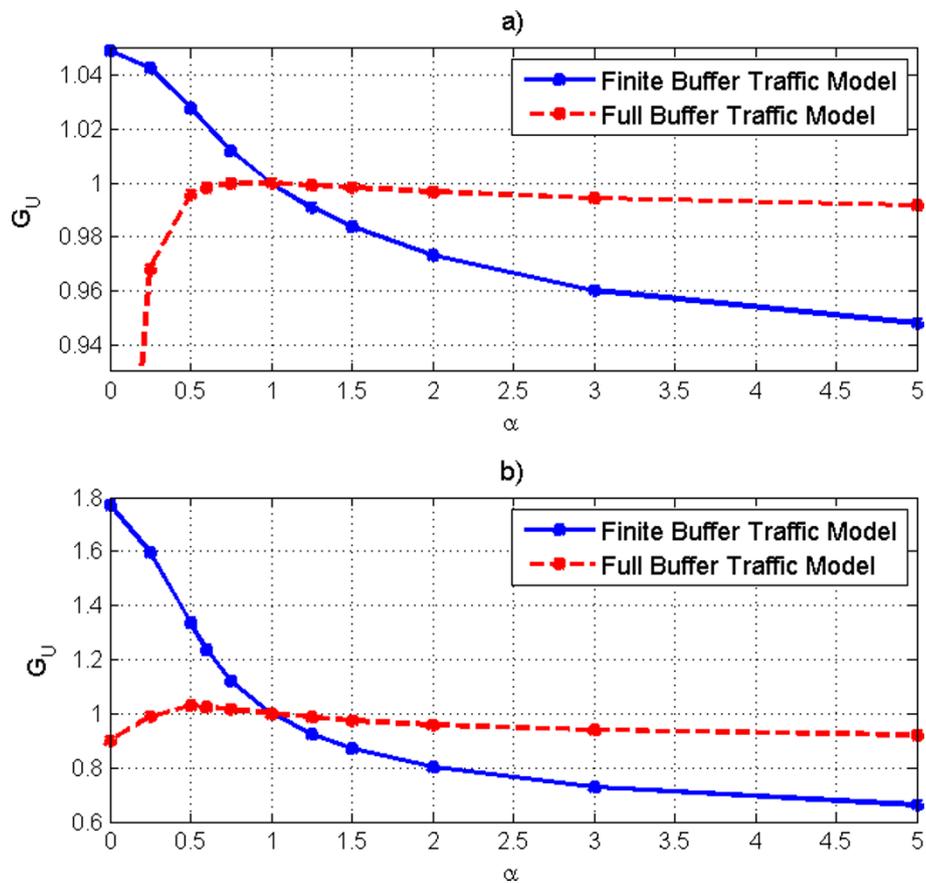
Figure 3 Average utility gain over Proportional Fair G_U for different α values under finite and full buffer traffic models. (a) Utility Function = $\log(r_i)$. (b) Utility Function = $2 \cdot \sqrt{r_i}$

performance of α = 0.6 and α = 1. For finite buffer, the offered load L = 11 Mbps is used.

Figure 5 shows the CDF curves of the user throughput for full and finite buffer models with the two schedulers. From the figure, it is seen that, with full buffer model, α = 0.6 prioritizes the users with good channel quality. As a consequence, 5% outage user throughput is significantly penalized. Meanwhile, in the case with the finite buffer model, α = 0.6 maintains approximately the same outage throughput and provides a gain in average user throughput. This behavior is similar to the one obtained in Section 5 with the simple OFDMA network model.

6.2. Performance results under different network load conditions

The performance with different cell loads is also evaluated under the 3G LTE network model. For the full buffer model, we vary the average number of users in the cell N = [8, 11, 14, 17, 20]; for the finite buffer model, the following offered cell loads are evaluated L = [8, 9, 10, 10.5, 11, 11.5] where L represents the offered cell load measured in Mbps.

The average user throughput gain over Proportional Fair G_Ω and user throughput at 5% outage gain over Proportional Fair G_O are presented for both traffic models in Figure 6. From the figure, it can be seen that the performance with full buffer is again approximately insensitive to the load level. When the finite buffer model is used, the relative performance of α = 0.6 improves with the cell load: for offered load below 10.5 Mbps there is small reduction of the outage but for loads above 10.5 Mbps α = 0.6 gains in outage user throughput; and regarding the average user throughput α = 0.6 provides a gain that reaches up to 60-70% for heavy offered load conditions (11-11.5 Mbps). This trend is again in agreement with the results presented in Section 5 for the simple OFDMA network model.

7. Conclusion

This article has studied the impact of the traffic model on the design of the OFDMA scheduling algorithms. Two simple traffic models proposed by the NGMN Alliance, the IEEE, and the 3GPP have been considered: the full buffer and the finite buffer models. With the first
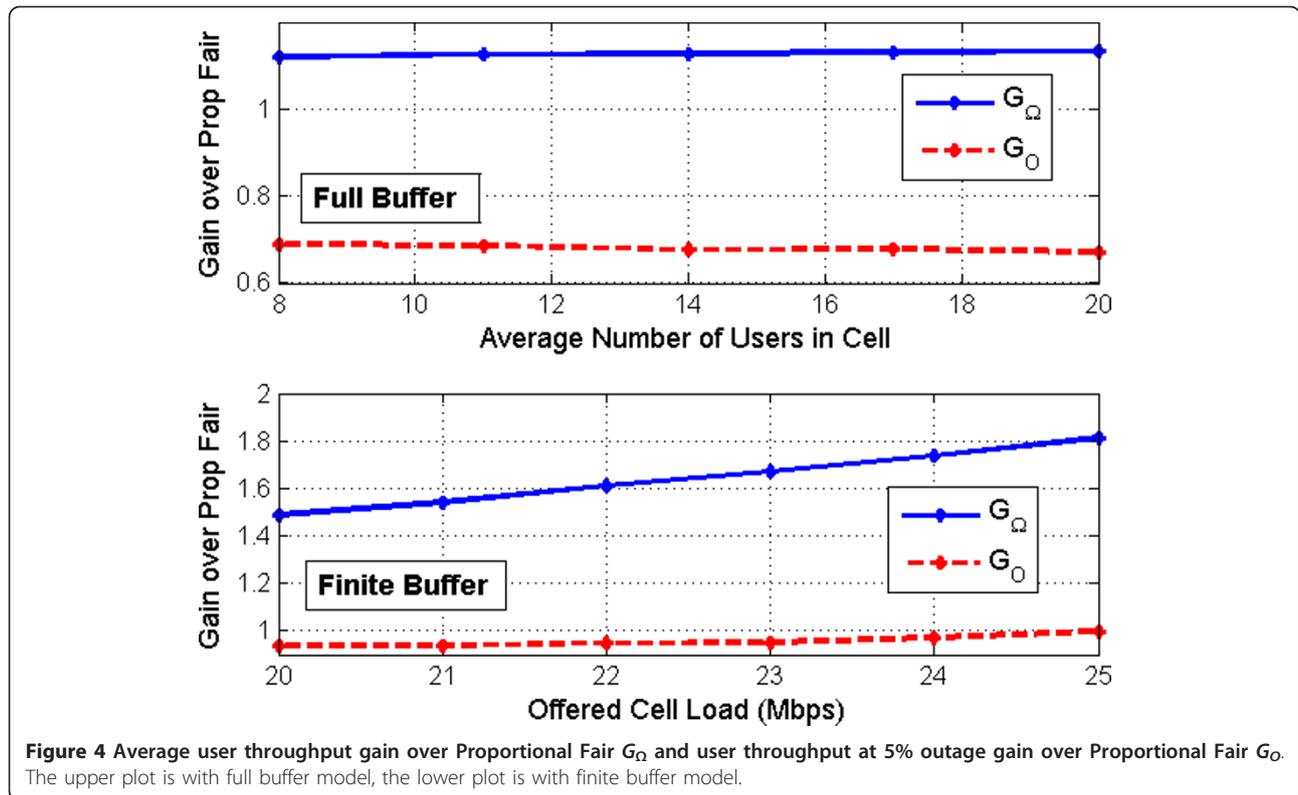
Figure 4 Average user throughput gain over Proportional Fair G_Ω and user throughput at 5% outage gain over Proportional Fair G_O. The upper plot is with full buffer model, the lower plot is with finite buffer model.

model, the number of users in the cell is constant and the buffers of the users' data flows always have unlimited amount of data to transmit. The second model includes a user birth process and every user is assigned a finite payload to transmit or receive.

In particular, this article has evaluated if the properties and trends of utility-based scheduling under the full

Table 2 3G LTE network model

Parameter	Setting/description
Number of sub-carriers per resource block	12
Pilot and control channel overhead	3/14 symbols
Modulation and coding rate settings	QPSK: 1/3, 1/2, 2/3 16QAM: 1/2, 2/3, 4/5 64QAM: 2/3, 4/5
HARQ model	Ideal chase combining
Number of SAW processes	6
CQI measurement sub band	360 kHz (2 RBs)
CQI measurement time window	1 ms
CQI reporting interval	5 ms
CQI feedback delay	2 ms
CQI quantization step	1.0 dB
Error in CQI estimation	Gaussian (zero mean, 1 dB std)
BLER target for Outer Loop LA	10%
Admission control constraint	Unlimited number of users per cell

buffer model also hold when the finite buffer is applied. Initially, a simple OFDMA network model has been used for the evaluation. The results have shown that the optimization objective of maximizing the aggregate utility through a gradient scheduling algorithm can only be achieved when the less realistic full buffer model is used. When the finite buffer model is applied, the gradient scheduling algorithm does not keep this property. In addition, under the full buffer model, the gradient scheduler exhibits a trade-off between average user throughput and the user throughput at 5% outage. However, under the more realistic finite buffer traffic model, the user throughput at 5% outage presents an optimum performance at around $\alpha = [0.6\text{-}1.0]$, and hence the gradient scheduler does not exhibit the mentioned trade-off. Moreover, the relative comparison of KPIs, such as the average user throughput or the user throughput at 5% outage, between versions of the α-fair gradient scheduling algorithm (with different fairness levels) is insensitive to the network load under the full buffer traffic model, but sensitive under the finite buffer traffic model. Furthermore, a realistic 3G LTE network model has also been used for the evaluation, where same trends are found as with the simple OFDMA network model.

Based on the presented findings, we propose as scheduling algorithm the gradient scheduler combined with
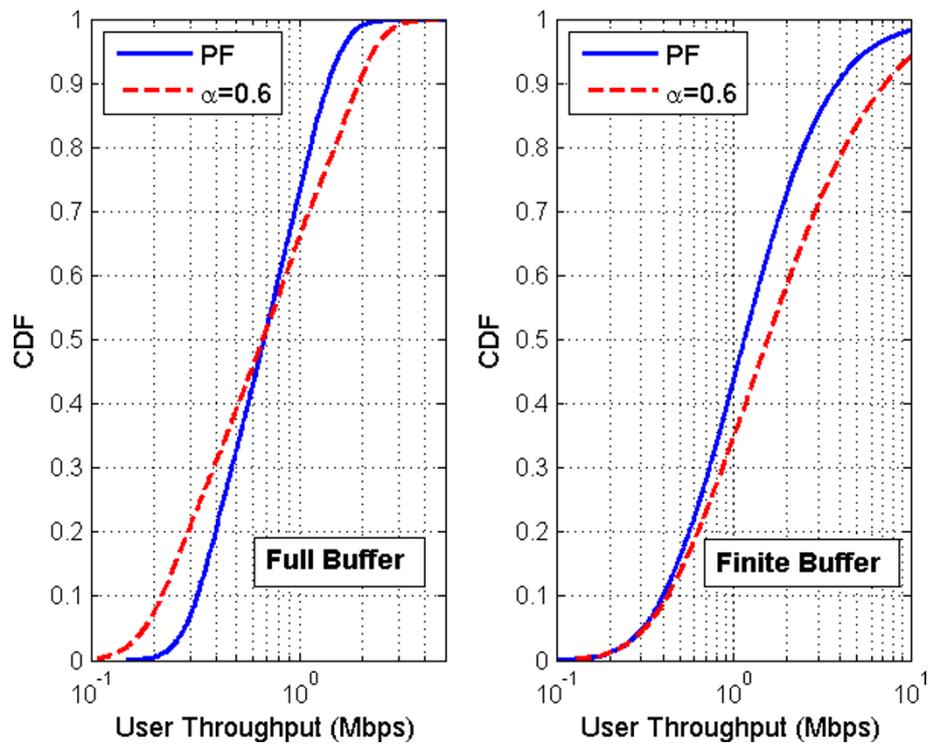
Figure 5 CDF of the user throughput for $\alpha = 0.6$ and $\alpha = 1.0$ (PF) under the full and finite buffer traffic models. 3G LTE network model.
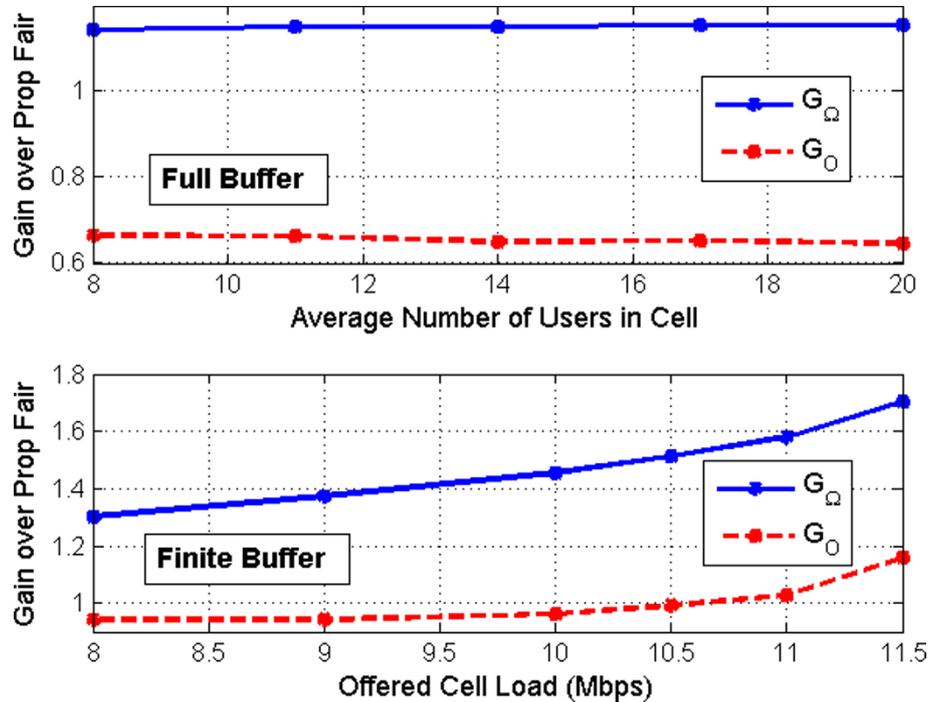

Figure 6 Average user throughput gain over Proportional Fair G_Ω and user throughput at 5% outage gain over Proportional Fair G_O. The upper plot is with full buffer model, the lower plot is with finite buffer model. 3G LTE network model.

the α-fair utility function and $\alpha = 0.6$ because it produces a similar user throughput at 5% outage, an average user throughput gain that reaches up to 60-70% for heavy offered load conditions, and an increased aggregate utility (assuming that the utility function is the Natural Logarithm) compared to the Proportional Fair algorithm. This performance gain could be considered in the implementation of QoS scheduling algorithms, which may benefit from incorporating this simple tuning.

These findings may also be an interesting trigger to additional investigations that further analyze if the presented trends also hold in other channel-unaware schedulers (e.g., Round Robin-based schedulers) in OFDMA systems.

Author details
[1]Department of Signal Theory, Telematics and Communications, University of Granada, Granada, Spain [2]Radio Access Technology Division, Department of Electronic Systems, Aalborg University, Aalborg, Denmark [3]Nokia Siemens Networks, Aalborg, Denmark

Competing interests
The authors declare that they have no competing interests.

References
1. 3GPP, Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description. Tech Spec 36.300 v10.5.0 (September 2011). http://www.3gpp.org
2. IEEE std, 802.16TM-2004, IEEE Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Fixed Broadband Wireless Access Systems (October 2004)
3. IEEE Std, 802.16e™-2005, IEEE Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems (February 2006)
4. 3GPP, Feasibility study for Further Advancements for E-UTRA (LTE-Advanced). Tech Rep 36.912 v10.0.0 (March 2011). http://www.3gpp.org
5. IEEE 802.16m-09/0034r4, IEEE 802.16m System Description Document (SDD) (December 2010) Available: http://www.ieee802.org
6. EC Strinati, G Corbellini, D Ktenas, HYGIENE scheduling for OFDMA wireless cellular networks, in *IEEE 69th Vehicular Technology Conference, 2009. VTC Spring 2009*, Barcelona, Spain, 1–5 (26-29 April 2009)
7. F Wang, A Ghosh, C Sankaran, P Fleming, F Hsieh, S Benes, Mobile WiMAX system: performance and evolution. IEEE Commun Mag. **46**(10), 41–49 (2008)
8. NGMN Alliance, Next Generation Mobile Networks Radio Access Performance Evaluation Methodology. White paper (January 2008). http://www.ngmn.org
9. IEEE 802.16 Broadband Wireless Access WG, IEEE 802.16m Evaluation Methodology Document (July 2008) http://www.ieee802.org
10. 3GPP, Further Advancements for E-UTRA Physical Layer Aspects. Tech Rep 36.814 v9.0.0 (March 2010). http://www.3gpp.org
11. Long Term Evolution (LTE), Overview of LTE Air-Interface. Technical White Paper (2007). http://www.motorola.com
12. N Wei, A Pokhariyal, TB Sorensen, T Kolding, P Mogensen, Performance of spatial division multiplexing MIMO with frequency domain packet scheduling: from theory to practice. IEEE Trans JSAC. **26**(6), 890–900 (2008)
13. Y Gao, Li Chen, X Zhang, Y Jiang, Performance evaluation of mobile WiMAX with dynamic overhead, in *Proc Of IEEE Vehicular Technology Conference (VTC), 2008* (September 2008)
14. E Rodrigues, F Casadevall, Resource allocation and packet scheduling in OFDMA-based cellular networks, in *4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications, 2009. CROWNCOM '09*, Hannover, Germany, 1–6 (22–24 June 2009)
15. J Gross, M Bohge, Dynamic mechanisms in OFDM wireless systems: a survey on mathematical and system engineering contributions. *TKN Technical Report* (May 2006)
16. T-D Nguyen, Y Han, A proportional fairness algorithm with QoS provision in downlink OFDMA systems. IEEE Commun Lett. **10**(11), 760–762 (2006)
17. G Song, Y Li, Cross-layer optimization for OFDM wireless network–Part I and Part II. IEEE Trans Wirel Commun. **4**(2), 614–634 (2005)
18. S Karpinski, EM Belding, KC Almeroth, Towards realistic models of wireless workload, in *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks and Workshops 2007. WiOpt*, Limassol, Cyprus, (2007)
19. G Divgi, E Chlebus, User and traffic characteristics of a commercial nationwide Wi-Fi hotspot network, in *IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007*, Athens, Greece, 1–5 (3–7 September 2007)
20. M Molina, P Castelli, G Foddis, Web traffic modeling exploiting TCP connections' temporal clustering through HTML-REDUCE. IEEE Netw. **14**, 46–55 (2001)
21. V Paxson, S Floyd, Wide area traffic: the failure of Poisson modeling. IEEE/ACM Trans Netw. **3**(3), 226–244 (1995). doi:10.1109/90.392383
22. K Johansson, J Bergman, D Gerstenberger, M Blomgren, A Wallen, Multi-carrier HSPA evolution, in *Proc of IEEE Vehicular Technology Conference (VTC), 2009*, Barcelona, Spain, 1–5 (April 2009)
23. SE Elayoubi, B Fourestie, On frequency allocation in 3G LTE systems, in *Proc of The 17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2006*, Helsinki, Finland, 1–5 (September 2006)
24. P Kela, J Puttonen, N Kolehmainen, T Ristaniemi, T Henttonen, M Moisio, Dynamic packet scheduling performance in UTRA long term evolution downlink, in *3rd International Symposium on Wireless Pervasive Computing, 2008. ISWPC 2008*, Santorini, 308–313 (7–9 May 2008)
25. C Wengerter, J Ohlhorst, AGE von Elbwart, Fairness and throughput analysis for generalized proportional fair frequency scheduling in OFDMA, in *Proc Of IEEE Vehicular Technology Conference(VTC), 2005*, vol. 3. Stockholm, Sweden, 1903–1907 (May 2005)
26. BG Lee, D Park, H Seo, *Wireless Communications Resource Management Wiley*, Hoboken, NJ, USA, (2009)
27. B Sadiq, R Madan, A Sampath, Downlink scheduling for multiclass traffic in LTE. EURASIP J. Wirel Commun Netw. **2009**, 1–18 (2009)
28. F Kelly, Charging and rate control for elastic traffic (corrected version). Eur Trans Telecommun. **8**(1), 33–37 (1997). doi:10.1002/ett.4460080106
29. A Pokhariyal, KI Pedersen, G Monghal, IZ Kovacs, C Rosa, TE Kolding, PE Mogensen, HARQ aware frequency domain packet scheduler with different degrees of fairness for the UTRAN long term evolution, in *IEEE 65th Vehicular Technology Conference, 2007. VTC2007-Spring*, Dublin, Ireland, 2761–2765 (April 2007)
30. H Holma, A Toskala, *LTE for UMTS–OFDMA and SC-FDMA Based Radio Access Wiley*, Hoboken, NJ, USA, (2009)
31. P Ameigeiras, JJ Ramos-Munoz, J Navarro-Ortiz, P Mogensen, JM Lopez-Soler, QoE oriented cross-layer design of a resource allocation algorithm in beyond 3G systems. Comput Commun J. **33**(5), 571–582 (2010). doi:10.1016/j.comcom.2009.10.016
32. G Song, Y Li, Utility-based resource allocation and scheduling in OFDM-based wireless broadband networks. IEEE Commun Mag. **43**(12), 127–134 (2005)
33. J Mo, J Walrand, Fair end-to-end window-based congestion control. IEEE/ACM Trans Netw. **8**(5), 556–567 (2000). doi:10.1109/90.879343
34. J Liu, A Proutiere, Y Yi, M Chiang, HV Poor, Stability, fairness, and performance: a flow-level study on nonconvex and time-varying rate regions. IEEE Trans Inf Theory. **55**(8), 3437–3456 (2009)
35. T Lan, D Kao, M Chiang, A Sabharwal, An axiomatic theory of fairness in network resource allocation. http://arxiv.org/abs/0906.0557v4 (October 2009)
36. PA Hosein, QoS control for WCDMA high speed packet data, in *4th International Workshop on Mobile and Wireless Communications Network, 2002*, Stockholm, Sweden, 169–173 (2002)

37. L Kleinrock, *Queuing Systems, Volume I: Theory Wiley*, Hoboken, NJ, USA, (1975)
38. A Pokhariyal, Downlink frequency-domain adaptation and scheduling–a case study based on the UTRA long term evolution. Ph D Thesis Dissertation (August 2007)
39. 3GPP, Deployment aspects. Tech Rep 25.943 v9.0.0 (December 2009). Available: http://www.3gpp.org
40. K Brueninghaus, D Astely, T Salzer, S Visuri, A Alexiou, S Karger, GA Seraji, Link performance models for system level simulations of broadband radio access systems, in *Proc IEEE PIMRC*, vol. 4. Berlin, Germany, 2306–2311 (September 2005)
41. KI Pedersen, G Monghal, IZ Kovacs, TE Kolding, A Pokhariyal, F Frederiksen, P Mogensen, Frequency domain scheduling for OFDMA with limited and noisy channel feedback, in *Proc of IEEE Vehicular Technology Conference(VTC), 2007*, Baltimore, MD, 1792–1796 (September 2007)