# Design of Fair Weights for Heterogeneous Traffic Scheduling in Multichannel Wireless Networks

Mehri Mehrjoo[1], Mohamad Khattar Awad[1], Mehrdad Dianati[2], and Xuemin (Sherman) Shen[1]

[1]Department of Electrical and Computer Engineering, University of Waterloo, Canada, N2L 3G1

[2]Faculty of Engineering and Physical Sciences, University of Surrey, UK, GU2 7XH

email:{mmehrjoo, xshen}@bbcr.uwaterloo.ca, mohamad@ieee.org, m.dianati@surrey.ac.uk

**Abstract**

Fair weights have been implemented to maintain fairness in recent resource allocation schemes. However, designing fair weights for multiservice wireless networks is not trivial because users' rate requirements are heterogeneous and their channel gains are variable. In this paper, we design fair weights for opportunistic scheduling of heterogeneous traffic in orthogonal frequency division multiple access (OFDMA) networks. The fair weights determine each user's share of rate for maintaining a utility notion of fairness. We then present a scheduling scheme which enforces users' long term average transmission rates to be proportional to the fair weights. Furthermore, the proposed scheduler takes the advantage of users' channel state information and flexibility of OFDMA resource allocation for efficient resource utilization. Simulation results demonstrate that the fair weights can be tuned for throughput-fairness trade off, and the scheduler can cope with the users movement or users' different resource requirements, while maintaining fairness.

## I. INTRODUCTION

Next-generation broadband wireless standards, such as IEEE 802.16, deploy orthogonal frequency division multiple access (OFDMA) mechanism to improve service provisioning and to overcome fading channel impairments. Deliberate resource scheduling of OFDMA networks facilitates different quality of service (QoS) provisioning, efficient utilization of limited resources available at the base station (BS), and maintaining fairness.

Pure opportunistic scheduling scheme allocates resources (i.e., sub-carriers, rate, power) to users with the highest channel gain. Although, the scheme is throughput-optimal [1], it results in unfair resource allocation [2]. In particular, such scheme may allocate all network resources to users with the highest channel gain and leaving the rest unsupported. Although this scheme maximizes the network throughput and hence the service providers' revenues, unsupported users may de-subscribe which reduces providers' revenues in the long run.

Recently, opportunistic fair scheduling schemes for multi-carrier networks have been appeared in the literature. In [3], an opportunistic fair scheduler for code division multiple access (CDMA) networks is proposed. The scheduling process is decoupled into a network throughput maximization process and a fairness process which can ensure probabilistic fairness or deterministic fairness. The probabilistic fairness maintains the differences among users' long-term throughput within a limited range with a bounded probability, and the deterministic fairness guarantees equal long-term throughput among users. However, it remains unclear how the expected differences are defined such that fairness is achieved. In addition, for heterogeneous traffic, maintaining equal long-term throughput among users may waste the resources of the network by allocating extra resources to a user. An opportunistic scheduler for OFDMA networks, which maintains temporal fairness or "utilitarian" fairness, is introduced in [4]. Whereas, in temporal fairness a certain long-term portion of time is allocated to each user, in "utilitarian" fairness a portion of the overall average throughput is allocated to each user. Temporal fairness criterion maintains resource fairness, i.e., resources are allocated on an equal time duration basis, which does not ensure performance (throughput) fairness in wireless networks. The portion of the overall average throughput that should be allocated to a user is pre-specified. A scheduling scheme based on high data rate (HDR) scheduling is presented for OFDMA networks in [5]. The HDR scheduling is a proportional fair scheduling scheme which has been proposed for scheduling data packets in *CDMA*$2000 1x$ evolution [6]. Although it attempts to reduce the complexity of the scheduling scheme by clustering the sub-carriers into sub-bands, a proof of proportional fairness is not given. Enforcing fairness through weighting factors is formulated in [7] for OFDMA networks. The focus is on proposing a linear complexity approach for sub-carrier assignment and rate allocations problems. The above scheduling schemes consider fairness provisioning among users with homogeneous rate requirements. However, the problem becomes complicated when users are heterogeneous in terms of traffic requirements and have non-concave utility functions. Resource scheduling for heterogeneous types of traffic needs to be revisited to optimally utilize resources while maintaining fairness among users. In addition, literature schemes address the complexity of the scheduling for multi-carrier networks assuming that the fair weights for different traffic types are known; however, finding proper weights is not trivial and remains an open problem.

In this paper, we design fair weights and accordingly propose a comprehensive solution for opportunistic fair scheduling of heterogeneous traffic. We design the fair weights for the users with different service rate (more precisely, utility) requirements, where the fair weights represent the fair proportions of the transmitted rates to users taking into account users' channel gain differences and heterogeneity of their data traffic. The fair weights are then used for opportunistic fair scheduling in the downlink of an OFDMA wireless network. The proposed scheduling scheme intends to maintain long-term fair allocation of sub-carriers and transmission power according to the weighting factors. In specific, we propose a modular scheduler consisting of a fairness module and a resource allocation module to separate fair weight computation from the resource allocation part. The fairness module executes a fairness scheme that generates a set of fair weights associated with users. These weights are periodically computed and fed to the resource allocation module so that resource allocation is simple and the required computing resources can be minimum. The resource allocation module allocates OFDMA sub-carriers and power based on users' instantaneous channel gains and fair weights. The fairness module performs in parallel with the resource allocation module, which allows for simple and fast scheduling. Numerical results show the proposed scheduling scheme is effective in maintaining fairness and achieving multi-user diversity gain for time variant channel and users with heterogeneous service demands.

The remainder of the paper is organized as follows. The system model and mathematical formulations of the OFDMA resource allocation and fairness modules are presented in section II. Formulated optimization problems for the scheduling are solved in section III. Numerical results are given in section IV, and the paper is concluded in section V.

## II. SYSTEM MODEL AND PROBLEM FORMULATIONS

The proposed scheduler consists of a fairness module and an OFDMA resource allocation module, as shown in Fig. 1. The notations $i$ and $j$ are user and sub-carrier indexes, respectively.

The fairness module includes the Fair Weight block, which computes the set of fair weights, $W_i$, and the Transmission History block, which computes the set of average transmitted rates, $R_i$. The Fair Weight block computes the fair weights, based on a fairness criteria, users' channel state information (CSI), and users' utilities. The Transmission History block updates the exponentially weighted moving average (EWMA) of transmitted rate to user $i$, $R_i$, at the beginning of each scheduling interval $n$ as given by

$$R_i(n) = (1 - \frac{1}{T_c})R_i(n-1) + (\frac{1}{T_c})r_i(n-1), \tag{1}$$

where $r_i$ is the transmitted rate to user $i$, and $T_c$ is a constant that determines rapidity of exponential decay. A larger $T_c$ results in rapid decay, and the converse is true. In addition, $T_c$ can be considered as the width of the moving average window so that when $T_c$ is large, averaging is performed over more numbers of scheduling intervals. EWMA emphasizes more on recent data by applying weighting factors that decrease exponentially. This technique is advantageous in the sense that the fairness scheme attempts to compensate for unfairness of recent allocations as much as possible.

The OFDMA resource allocation module determines allocated rates to users based on $a_{ij}$ and $\frac{R_i}{W_i}$. Users channel diversity gain is achieved by taking into account users' CSI, $a_{ij}$, and fairness is compensated for by considering $\frac{R_i}{W_i}$ as a measure of fairness deficit. When $\frac{R_i}{W_i}$ is close to 1,

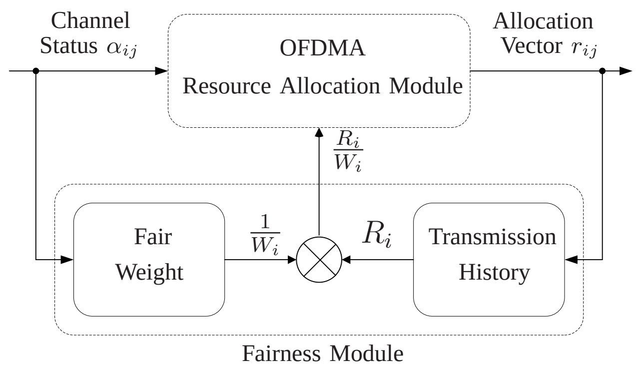


Fig. 1. Architecture of proposed scheduler

the average transmitted rate to user $i$ has been fair enough in the past scheduling intervals. On the other hand, $\frac{R_i}{W_i} << 1$ or $\frac{R_i}{W_i} >> 1$ mean starvation and overallocation of user $i$, respectively, where the scheduler should compensate for that in the next scheduling intervals.

We formulate the OFDMA resource allocation problem and the fair weight design problem separately using optimization programming techniques. The OFDMA resource allocation, described in subsection II-A, is an optimization problem where its objective function and constraints model the scheduling scheme and OFDMA specifications. Similarly, we present an optimization problem that considers users' heterogeneous rate requirements and CSI to compute proportional fair weights in subsection II-B.

### *A. OFDMA Resource Allocation*

We consider the downlink of a single BS and multiple users located in one hop neighborhood from the BS. Users' backlogged traffic, buffered in separate queues at the BS, is scheduled at the beginning of each downlink frame consisting of $N_s$ OFDM symbols. The BS assigns OFDM sub-carriers to users and allocates a fraction of its power, $P_{BS}$, to each sub-carrier at each scheduling instance. Relevant system parameters in our model are defined in Table I.

Without loss of generality, we assume that the spectral density of noise and sub-carriers bandwidth are equal to one. Thus, the allocated rate to user $i$ on sub-carrier $j$ of symbol $n$,

Table I
LIST OF SYMBOLS

| Symbol | Description |
|---|---|
| $N_u$ | number of users in the network |
| $N_c$ | number of OFDMA sub-carriers |
| $N_s$ | number of OFDMA symbols in the down-link frame |
| $N_f$ | number of OFDMA symbols in fair weights computations |
| $i$ | user index belongs to $\mathcal{N}_u := \{1, 2, ..., N_u\}$ |
| $j$ | sub-carrier index belongs to $\mathcal{N}_c := \{1, 2, ..., N_c\}$ |
| $n$ | symbol index belongs to $\mathcal{N}_s := \{1, 2, ..., N_s\}$ |
| $R_i$ | average transmitted rate to user $i$ |
| $W_i$ | fair weight of user $i$ |
| $P_{BS}$ | the BS total power budget |
| $\alpha_{ijn}$ | channel gain of user $i$ on sub-carrier $j$ of OFDMA symbol $n$ |
| $p_{ijn}$ | required power by user $i$ on sub-carrier $j$ of OFDMA symbol $n$ to transmit $r_{ijn}$ |
| $r_{ijn}$ | achievable rate by user $i$ on sub-carrier $j$ of OFDMA symbol $n$ |

denoted by $r_{ijn}$, is

$$r_{ijn} = \log_2 (1 + \alpha_{ijn} p_{ijn}) \,. \tag{2}$$

Total allocated power to the sub-carriers of each OFDMA symbol is limited by $P_{BS}$, i.e.,

$$\sum_{i=1}^{N_u} \sum_{j=1}^{N_c} p_{ijn} \leq P_{BS} \quad \forall n \in \mathcal{N}_s. \tag{3}$$

Implementation of OFDMA requires exclusive allocation of a sub-carrier to a single user. This constraint can be represented by

$$r_{\hat{i}jn} \cdot r_{ijn} = 0 \quad \forall \hat{i} \in \mathcal{N}_u, i \neq \hat{i}, \;\; \forall j \in \mathcal{N}_c, \;\; \forall n \in \mathcal{N}_s. \tag{4}$$

Constraint (4) implies that if sub-carrier $j$ is assigned to user $\hat{i}$, i.e., $r_{\hat{i}jn} \neq 0$, the allocated rate to every other user on sub-carrier $j$ of OFDMA symbol $n$ must be zero.

The trade-off between throughput and fairness can be tuned by the objective function. As different notions of fairness in long or short term basis are expected in practice, the objective function is defined in such a way to be adjusted for different fairness demands. We consider $\frac{R_i}{W_i}$ as the fairness tuning term in the objective function. Recall that $W_i$ is a fair allocation rate to user $i$, and $R_i$ is the window average of transmitted rate to user $i$. Depending on how long the average window length of $R_i$ is, the scheduler reaction time to the unfairness changes. Upon short average window length, $R_i$ approaches zero very fast, if user $i$ does not receive any rate for multiple scheduling intervals in a row. In other words, a short average window length forces the scheduler to short term fairness, so the scheduler rate allocation is expected to be in favor of fairness provisioning than throughput maximization. On the other hand, a long average window length for $R_i$ allows the scheduler to compensate for the fairness in longer time and take advantage of users' channel diversity to improve throughput. Therefore, to catch the users channel diversity gain into account and maintain fairness simultaneously, the objective function is written as

$$\max \sum_{n=1}^{N_s} \sum_{j=1}^{K} \sum_{i=1}^{M} \left( \frac{r_{ijn}}{\frac{R_i}{W_i}} \right). \tag{5}$$

Accordingly, the probability of assigning sub-carrier $j$ to user $i$ increases when the achievable transmission rate of user $i$ on sub-carrier $j$ is high or the average transmitted rate to user $i$ is smaller than its fair weight. Different degrees of performance trade-off between throughput and fairness can be obtained and optimized [8], which is out of the scope of this paper.

The objective function (5) along with constraints (2), (3), (4), represent opportunistic fair scheduling by an optimization problem, denoted by $(P1)$ as follows.

$$P_1: \quad \max_{r_{ijn}} \sum_{n=1}^{N_s} \sum_{j=1}^{N_c} \sum_{i=1}^{N_u} \left( \frac{r_{ijn}}{\frac{R_i}{W_i}} \right) \tag{6}$$

$$\text{s.t} \quad \sum_{i=1}^{N_u} \sum_{j=1}^{N_c} \frac{2^{r_{ijn}} - 1}{\alpha_{ijn}} \leq P_{BS} \; \forall n \in \mathcal{N}_s, \tag{7}$$

$$r_{\hat{i}jn} \cdot r_{ijn} = 0 \quad \forall \hat{i} \in \mathcal{N}_u, i \neq \hat{i}, \; \forall j \in \mathcal{N}_c, \; \forall n \in \mathcal{N}_s, \tag{8}$$

$$r_{ijn} \geq 0 \quad \forall i \in \mathcal{N}_u, \; \forall j \in \mathcal{N}_c, \; \forall n \in \mathcal{N}_s. \tag{9}$$

Problem $P_1$ is solved in each scheduling interval to obtain allocated rate to users on all sub-carriers for each OFDMA symbol. Since CSI may not be received accurately at the BS, we have proposed an approach in [9] to account for CSI inaccuracy resulted from estimation error and feedback delay. In practice, providing CSI of each sub-carrier over all symbols of each scheduling interval results in large messaging overhead on the reverse feedback channel. Because of the correlation among CSI of a sub-carrier over consecutive symbols, the CSI of each sub-carrier is assumed to be constant for all symbols over a scheduling interval. Accordingly, index $n$ representing symbols of each scheduling interval can be dropped; thus, P1 can be reduced to a simpler optimization problem, denoted by $P_2$:

$$P_2: \quad \max_{r_{ij}} \sum_{j=1}^{N_c} \sum_{i=1}^{N_u} \left( \frac{r_{ij}}{\frac{R_i}{W_i}} \right) \tag{10}$$

$$\text{s.t} \quad \sum_{i=1}^{N_u} \sum_{j=1}^{N_c} \frac{2^{r_{ij}} - 1}{\alpha_{ij}} \leq P_{BS}, \tag{11}$$

$$r_{\hat{i}j} \cdot r_{ij} = 0 \quad \forall \hat{i} \in \mathcal{N}_u, i \neq \hat{i}, \ \forall j \in \mathcal{N}_c, \tag{12}$$

$$r_{ij} \geq 0 \quad \forall i \in \mathcal{N}_u, \ \forall j \in \mathcal{N}_c, \tag{13}$$

where $r_{ij}$ represents allocated rate to user $i$ on sub-carrier $j$ of all symbols in each scheduling interval.

### *B. Fair Weight Design*

Several factors should be taken into account for designing fair weights. First, users' channel status is diverse, so the resource allocation should be updated frequently. Second, users' resource requirements depends on the service they receive. Third, fairness criteria are not unique, and different fairness criteria, such as *proportional*, $\alpha-$*fair*, or *maxmin* fairness [10], [11] can be applied to users' rates or utilities.

Wireless channel suffers from fast and slow channel variations. Fast variations are highly unpredictable, so they are not considered in a long term resource allocation scheme. We look at the long trend of wireless channel, happened over multiple frames, and design the fair weights

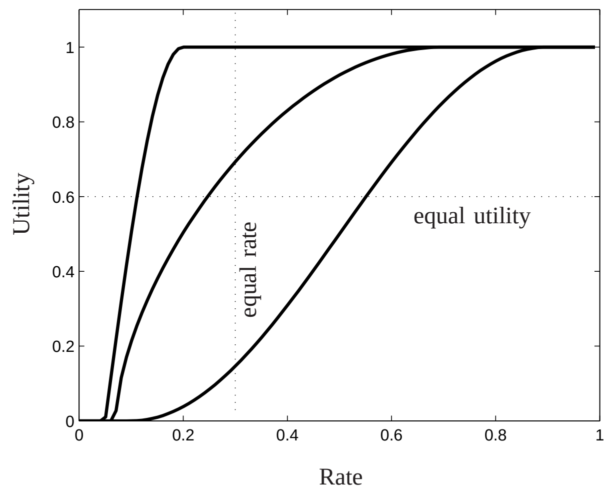


Fig. 2. Comparison between equal rate and equal utility allocation

accordingly. These fair weights are used in resource allocation for the next multiple frames. However, to adapt to the channel variations, the fair weights are updated periodically, dependent to channel statistics. To allocate resources based on the application requirements of users, utility-based allocation can be employed. Fig. 2 shows the utilities of three different applications. The dotted line, labeled "equal rate", illustrates that equal rate allocation does not provide equal user satisfaction. On the other hand, equal allocation of utilities, which is interpreted as equal users' satisfaction, utilizes the network resources more efficiently [12]. Thus, we consider utility fairness instead of rate fairness.

The fair weights are determined based on utility proportional fairness where the allocated resources are proportional to users' demands. Consider $\mathcal{U} = \{\boldsymbol{U_k} | \boldsymbol{U_k} = \{u_{k1}, u_{k2}, \ldots, u_{kN_u}\}\}$, a bounded set of $N_u$ users' feasible utility subset $\boldsymbol{U_k}$, where $u_{ki}$ is the user $i$'s utility. Utility proportional fairness is defined as follows [13].

*Definition 2.1:* A set of utilities $\boldsymbol{U_x} = \{u_{x1}, u_{x2}, \ldots, u_{xN_u}\}$ is *utility proportional fair* if for any feasible utility set $\boldsymbol{U_y} = \{u_{y1}, u_{y2}, \ldots, u_{yN_u}\}$, the sum of proportional changes in their utilities is non-positive, i.e.,

$$\sum_{i=1}^{N_u} \frac{u_{yi}(r_{yi}) - u_{xi}(r_{xi})}{u_{xi}(r_{xi})} \leq 0. \tag{14}$$

Note that $u_{ki}$ is a function of allocated rate. Therefore, $\boldsymbol{U_k}$ is utility proportional fair if the set of rates $\{r_{k1}, r_{k2}, \ldots, r_{kN_u}\}$ is found to satisfy (14). A straightforward way to obtain a proportional fair allocation $\boldsymbol{U_k} \in \mathcal{U}$ is to maximize $\sum_i \log(u_{ki})$ over the convex set of feasible allocations $\mathcal{U}$:

$$\max_{k} \ \ \mathcal{F} = \sum_{i} \log(u_{ki}). \tag{15}$$

Accordingly, a rate allocation, denoted by $\{w_{ijn}\}$, which is utility proportional fair can be obtained by solving the following optimization problem:

$$P_3: \ \max_{w_{ijn}} \ \mathcal{F} \tag{16}$$

$$\text{s.t} \quad \sum_{i=1}^{N_u} \sum_{j=1}^{N_c} \frac{2^{w_{ijn}} - 1}{a_{ijn}} \leq P_{BS} \ \forall n \in \{1, 2, \cdots, N_f\}, \tag{17}$$

$$w_{ijn} \geq 0 \quad \forall i \in \mathcal{N}_u, \quad \forall j \in \mathcal{N}_c, \quad \forall n \in \{1, 2, \cdots, N_f\}. \tag{18}$$

Problem $P_3$ has a power constraint similar to (2) and (3). Also, the exclusive sub-carrier assignment restriction, constraint (4), is relaxed because this problem is solved for fair weights regardless of specific sub-carrier assignments. The fair rate allocation is computed over $N_f$ symbols, where $N_f$ contains multiple frames, i.e., $N_f >> N_s$. As fairness is based on large scale variations of users' channel, $w_{ijn}$ are summed up over $N_f$ and $N_c$. Therefore, the proportional fair rate allocation to user $i$, i.e., the user $i$'s fair weight is:

$$W_i = \sum_{j=1}^{N_c} \sum_{n=1}^{N_f} w_{ijn}. \tag{19}$$

The fair weights represent the rate fractions that users should receive over a long time with respect to other users, i.e., $\sum_{i \in \mathcal{N}_u} W_i = 1$. The OFDMA resource allocation module may allocate more or less rate than the fair rate to each user in each scheduling instance. However, it attempts to maintain the following equalities over a long time [14]:

$$\frac{R_1}{W_1} = \frac{R_2}{W_2} = \cdots = \frac{R_{N_u}}{W_{N_u}}. \tag{20}$$

If the scheduler allocates the available resources to users such that the set of aggregate transmitted rates to users is proportional to the set of fair weights, $W_i$, i.e., equation (20) is satisfied, the scheduling scheme is utility proportional fair.

## III. OFDMA RESOURCE ALLOCATION AND FAIR WEIGHT DESIGN SOLUTIONS

Problems $P_2$ and $P_3$ need to be solved in every scheduling interval. Problems $P_2$ and $P_3$ are non-convex optimization problems in general, and finding their optimal solutions is nontrivial [15]. Problem $P_2$ is non-convex because of its discrete feasible region, while $P_3$ is non-convex because of the non-convex utility functions in the objective function. The efficiency in solving a non-convex problem strongly depends on how non-convexity of the problem is treated. Therefore, we apply two different approaches to treat the non-convexity of each problem:

- First, we use a Lagrange dual decomposition method to solve $P_2$. The method does not guarantee an optimal solution, but it can efficiently obtain near optimal solution(s) with a practical number of sub-carriers [16]. The adaptation of Lagrange dual decomposition method hinges on the results reported in [17] that the duality gap[1] vanishes as the number of sub-carriers increases.
- Second, an interior point method is applied to solve $P_3$ because the objective function is the sum of users' utilities which can be non-linear functions of users' rates, and interior point methods can efficiently solve non-linear optimization problems [18].

### A. *Solution of the OFDMA Resource Allocation Problem* $P_2$

If $\delta_i = W_i/R_i$, the objective function of problem $P_2$ is a maximization of $\sum_{i=1}^{N_u}\left(\delta_i \sum_{j=1}^{N_c} r_{ij}\right)$. Constraints (12) and (13) form the domain $\mathcal{D}$ over which the Lagrangian of $P_2$ can be defined

[1]The difference between the primal optimal and dual optimal solutions

as

$$\mathcal{L}\left(\{r_{ij}\},\lambda\right) = \sum_{i=1}^{N_u}\sum_{j=1}^{N_c}\delta_i r_{ij} - \lambda\left(\frac{2^{r_{ij}}-1}{\alpha_{ij}} - P_{BS}\right), \tag{21}$$

where $\lambda$ is the Lagrange multiplier. The dual problem of $P_2$, can be expressed as

$$\min_{\lambda}\max_{\{r_{ij}\}\in\mathcal{D}}\mathcal{L}\left(\{r_{ij}\},\lambda\right). \tag{22}$$

From the solution of the dual problem, the set of rate allocations $r_{ij}$, $\forall i$ & $j$, can be determined. The optimization problem (22) is a minimization problem with one scalar variable $\lambda$ that can be solved by an iterative algorithm (Algorithm 1). In each iteration of Algorithm 1, the set of $r_{ij}$ that maximizes $\mathcal{L}$ is determined by solving $N_c$ decomposed problems of rate allocation to sub-carriers. As allocation of sub-carriers to users are independent, the optimization problems (23) can be solved in parallel to obtain allocated rate to sub-carriers.

$$\max_{\{r_{ij}\}\in\mathcal{D}}\sum_{i=1}^{N_u}\delta_i r_{ij} - \lambda\left(\frac{2^{r_{ij}}-1}{\alpha_{ij}}\right)\ \forall j = 1\cdots N_c. \tag{23}$$

When adaptive modulation is used, allocated number of bits to each sub-carrier is a discrete variable that can be chosen from the bit loading vector of the modulation technique [19]. Accordingly, the solution of problem (23) is determined by searching over the domain $\mathcal{D}$. The search is performed in real-time because the size of the domain $\mathcal{D}$ is confined by the number of modulation levels, users, and sub-carriers.

### *B. Solution of the Fair Weight Design Problem* $P_3$

For notational simplicity, a solution of $P_3$ is denoted by a rate allocation vector **w**:

$$\mathbf{w} = [w_{11}, w_{12}, \ldots, w_{1K}, \ldots, w_{M1}, \ldots, w_{MK}]^T, \tag{24}$$

where $w_{ij}$ represents allocated rate to user $i$ on sub-carrier $j$ and $w_i = \sum_{j=1}^{N_c} w_{ij}$ is allocated rate to user $i$. We form a vector $\mathbf{c}(\mathbf{w})$ of the inequality constraints (17) and (18), and convert the inequality constraints to equality constraints by associating a positive slack variable to each

**Algorithm 1** Solution algorithm for the dual problem of $P_2$

**Input**: $N_u,\ N_c,\ P_{BS},\ \alpha_{ij},\ \delta_i,\ bit\ loading\ set$
**Result**: $r_{ij}$
**begin**
  **Setting up and initialization:**
  Set $h = 1$, $\epsilon = 1$, $Exit\_flag = 1$, $\lambda_{h-1} = \lambda_h = 0$.
  Solve (23) for $r_{ij}$.
  Compute $\Delta p = P_{BS} - p_{ij}$.
  **if** $\Delta p > 0$ **then**
    **return** $r_{ij}$.
  **else**
    **while** $Exit\_flag > 1e-5$ **do**
      **if** $\Delta p > 0$ **then**
        $\epsilon = 0.99 * \epsilon$.
        $\lambda_h = \lambda_{h-1}$.
        $\Delta p_h = \Delta p_{h-1}$.
      **else**
        $\lambda_{h-1} = \lambda_h$.
        $\Delta p_{h-1} = \Delta p_h$.
      **end**
      $\lambda_h = \lambda_h + |\epsilon * \Delta p|$.
      Solve (23) for $r_{ij}$.
      Update $\Delta p$.
      $Exit\_flag = \lambda_h - \lambda_{h-1}$.
      $h = h + 1$.
    **end**
  **end**
  **return** $r_{ij}$.
**end**

constraint. Denote the $(2M+1)N_c$ vector of slack variables by **s**. Hence, $P_3$ is converted to the following minimization problem:

$$P_4: \quad \min_{\mathbf{w}} \quad -\sum_i log(u_{ki}(\mathbf{w})) \tag{25}$$

$$\text{s.t} \quad \mathbf{c}(\mathbf{w}) - \mathbf{s} = 0, \tag{26}$$

$$\mathbf{s} \geq 0. \tag{27}$$

To find an approximation for a local optimum of a nonlinear problem, the interior point algorithm solves a series of perturbed Karush-Kuhn-Tucker (KKT) conditions of $P_4$:

$$\nabla \mathbf{u}(\mathbf{w}) - \mathbf{A}^T(\mathbf{w})\mathbf{z} = 0, \tag{28}$$

$$\mathbf{c}(\mathbf{w}) - \mathbf{s} = 0, \tag{29}$$

$$\mathbf{S}\mathbf{z} = \mu \mathbf{e}, \tag{30}$$

$$\mathbf{s} \geq 0, \quad \mathbf{z} \geq 0\ , \tag{31}$$

with $\mathbf{e} = (1, 1, ..., 1)^T$ and $\mu > 0$.

In the perturbed KKT conditions, $\mathbf{S}$ is a diagonal matrix with diagonal elements given by vector $\mathbf{s}$, and vector $\mathbf{z}$ contains $(2M + 1)N_c$ Lagrange multipliers used in the definition of the Lagrangian function of $P_4$:

$$\mathcal{L}\left(\mathbf{w}, \mathbf{s}, \mathbf{z}\right) = log(\mathbf{u}(\mathbf{w})) - \mathbf{z}^T \left(\mathbf{c}(\mathbf{w}) - \mathbf{s}\right). \tag{32}$$

The matrix $\mathbf{A}$ in (28) is the Jacobian matrix of $\mathbf{c}(\mathbf{w})$.

Interior point methods begin with an initial interior point in the feasible region that satisfies perturbed KKT conditions for some $\mu$ and proceeds to find another interior point that satisfies perturbed KKT conditions for a smaller value of $\mu$. As the algorithm evolves, $\mu$ decreases, and consequently the solution of the perturbed KKT conditions approaches the solution of the KKT conditions, where $\mu = 0$. It is expected that after several iterations the solution will converge to a point that satisfies the KKT conditions of the problem [18].

In each iteration of the interior point method, the directions and lengths of steps from one interior point to another are updated based on the first and second order gradients of objective function and constraints. At each iteration, step direction for each of the variables $\mathbf{w}$, $\mathbf{s}$, and $\mathbf{z}$, i.e., $\mathbf{b} = [\mathbf{b_w}, \mathbf{b_s}, \mathbf{b_z}]^T$, are computed by solving the following linear system of equations:

$$\begin{pmatrix} \nabla^2_{\mathbf{ww}}\mathcal{L} & 0 & -\mathbf{A}^T(\mathbf{w}) \\ 0 & \mathbf{Z} & \mathbf{S} \\ \mathbf{A}(\mathbf{w}) & -\mathbf{I} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{b_w} \\ \mathbf{b_s} \\ \mathbf{b_z} \end{pmatrix} = \begin{pmatrix} \nabla_{\mathbf{w}}\mathbf{u}(\mathbf{w}) - \mathbf{A}^T(\mathbf{w})\mathbf{z} \\ \mathbf{Sz} - \mu\mathbf{e} \\ \mathbf{c(w)} - \mathbf{s} \end{pmatrix}, \tag{33}$$

Here, $\mathbf{Z}$ denotes the diagonal matrix whose diagonal elements are given by vector $\mathbf{z}$.

After obtaining step directions, the length of step in each direction, step length, denoted with

$\alpha_s^{max}$ and $\alpha_z^{max}$, are specified as:

$$\boldsymbol{\alpha}_{\mathbf{s}}^{max} = \max\left\{\alpha \in (0,1] : \mathbf{s} + \alpha \mathbf{b}_{\mathbf{s}} \geq (1-\tau)\,\mathbf{s}\right\}, \tag{34}$$

$$\boldsymbol{\alpha}_{\mathbf{z}}^{max} = \max\left\{\alpha \in (0,1] : \mathbf{z} + \alpha \mathbf{b}_{\mathbf{z}} \geq (1-\tau)\,\mathbf{z}\right\}, \tag{35}$$

where $\tau \in (0,1)$. A small value of $\tau$ forces **s** and **z** to approach zero very quickly, so a large value of $\tau$ close to one, e.g., $\tau = 0.995$, is usually chosen. The new interior point, slack variables, and Lagrange multipliers, $(\mathbf{w}^+, \mathbf{s}^+, \mathbf{z}^+)$, are determined with the information of step directions and step lengths accordingly:

$$\mathbf{w}^+ \;=\; \mathbf{w} + \boldsymbol{\alpha}_s^{max}\mathbf{b}_{\mathbf{w}}, \tag{36}$$

$$\mathbf{s}^+ \;=\; \mathbf{s} + \boldsymbol{\alpha}_s^{max}\mathbf{b}_{\mathbf{s}}, \tag{37}$$

$$\mathbf{z}^+ \;=\; \mathbf{z} + \boldsymbol{\alpha}_z^{max}\mathbf{b}_{\mathbf{z}}. \tag{38}$$

For the next iteration, $\mu$ is updated to a smaller value, $\mu^+ < \mu$, via a linear method:

$$\mu^+ = \sigma\mu \qquad \sigma \in (0,1). \tag{39}$$

Since $\sigma < 1$, $\mu$ approaches to zero over several iterations. However, choosing a very small $\sigma$ or a very large $\sigma$ will cause faster or slower convergence, respectively. Although fast convergence is always desired, it may force some algorithm parameters, such as **s** and **z**, to approach zero too quickly, which degrades the performance of the algorithm, e.g., the offered solution may be infeasible or far from optimality.

The interior point algorithm is terminated when a stopping criterion is satisfied. In this work, an initial value of $\mu_0 = 1$ has been chosen, and when $\mu$ approaches a very small value or the change in allocated weight vector, **w**, is negligible, the algorithm stops. Algorithm 2 presents a summary of the interior point algorithm used in our simulation.

**Algorithm 2** The interior point algorithm for $P_4$

**Input**: $N_u,\ N_c,\ P_{BS},\ B,\ \alpha,$ *users' utilities*, $\boldsymbol{w_{initial}},\ \boldsymbol{s_0},\ \mu_0,\ \tau,\ \sigma$
**Result**: **w**
**begin**
  **Setting up and initialization:**
  Choose $\boldsymbol{w_{initial}}$ and compute $\boldsymbol{s_0} > 0$.
  Choose $\mu_0 > 0$ and compute $\boldsymbol{z_0} > 0$ accordingly.
  Set parameters $\tau \in (0,1)$ and $\sigma \in (0,1)$.
  Set $k = 0$.
  **while** $Exit\_flag == 0$ **do**
    Solve (33) to obtain step direction $\mathbf{b} = [\mathbf{b_w}, \mathbf{b_s}, \mathbf{b_z}]^T$.
    Compute $\boldsymbol{\alpha}_{\mathbf{s}}^{max}$, and $\boldsymbol{\alpha}_{\mathbf{z}}^{max}$ using (34) and (35).
    Compute $(\mathbf{w}^{k+1}, \mathbf{s}^{k+1}, \mathbf{z}^{k+1})$ using (36) to (38).
    Set $\mu^{k+1} \leftarrow \mu^k$ and $k \leftarrow k+1$.
    Compute $Exit\_flag$.
  **end**
  **return w**.
**end**

### *C. Complexity of Proposed Approach*

The decomposition of (22) into $N_c$ equations (23) reduces the problem's exponential complexity to a linear one in terms of $N_c$ [17]. The solution of (23) is obtained by a heuristic search method because of the non-convexity of the domain $\mathcal{D}$. The size of $\mathcal{D}$ is confined by the number of modulation levels, users, and sub-carriers, denoted by $Q$, $N_u$, and $N_c$, respectively. In each iteration of the $while$ loop in Algorithm 1, a set with $QM$ size is searched for rate allocation to each sub-carrier. If the $while$ loop requires $N_{while}$ iterations to converge, then the set of equations (23) will be solved in $KQMN_{while}$ iterations. Whereas solving the equation (22) with an exhaustive search requires searching over a set of size $(QM)_c^N$.

Problem $P_3$ is required to be solved only when the network characteristics, such as users' average channel gains or the number of admitted users to the network, change. The scheduling scheme starts with default fair weights, e.g., all equal to one, and updates the fair weights with the ones obtained by solving $P_3$ during the first iteration of the scheduling scheme.

## IV. NUMERICAL RESULTS

Performance of the opportunistic fair scheduling scheme is evaluated in this section. Performance metrics are throughput and fairness index which are compared with those of a pure opportunistic scheduling scheme.

To compare the performance in terms of fairness, a fairness metric needs to be defined first. Gini fairness index, which is an inequality measure of resource sharing, measures deviation from equations (20) for each scheduler. Let the total allocated rate to user $i$ over the simulated intervals be symbolized by $\tilde{R}_i$. We examine the inequality among the set of proportions $\mathbf{v} = \{v_i \mid v_i = \tilde{R}_i/W_i\}$ by Gini fairness index, $G_{FI}$, defined as follows:

$$G_{FI} = \tfrac{1}{2M^2\bar{\mathbf{v}}} \sum_{x=1}^{N_u} \sum_{y=1}^{N_u} |\mathbf{v}_x - \mathbf{v}_y|, \tag{40}$$

where

$$\bar{\mathbf{v}} = \frac{\sum_{i=1}^{N_u} \mathbf{v}_i}{N_u}. \tag{41}$$

The Gini fairness index takes a value between 0 and 1. A rate allocation is perfectly fair if $G_{FI} = 0$. A high value of $G_{FI}$ indicates higher unfairness among the proportions.

The wireless channel is simulated to experience both frequency selective and large-scale fading [20], [21]. The users receive six Rayleigh distributed multipath signals. The real and imaginary components of the received signals to different users are generated from an uncorrelated multidimensional Gaussian distribution with zero mean and an identity covariance matrix. The large-scale fading is distance dependent and follows the inverse-power law [20]:

$$|\gamma_{ij}|^2 = D_i^{-\kappa} |\alpha_{ij}|, \tag{42}$$

where $D_i$ is the distance between the BS and user $i$ in meters, $\kappa$ is path loss exponent, and $\gamma_{ij}$ is path loss of user $i$ on sub-carrier $j$. The numerical values of the wireless channel used in the simulation are: Doppler frequency$=30$ Hz, and $\kappa = 2$.

The network supports users with non-concave and concave utility functions, respectively. The users' utility functions are expressed by equation (43) [22], where **r** denotes allocated rate to the user, $l1$ and $l2$ are lower and upper rate thresholds, and $k$ controls the convexity of the utility function. The function is concave for $k < 1$ and convex for $k > 1$. $(k = 0.7,\ l1 = 1,\ l2 = 800)$ and $(k = 2,\ l1 = 10,\ l2 = 600)$ have been chosen for concave and non-concave utility functions,

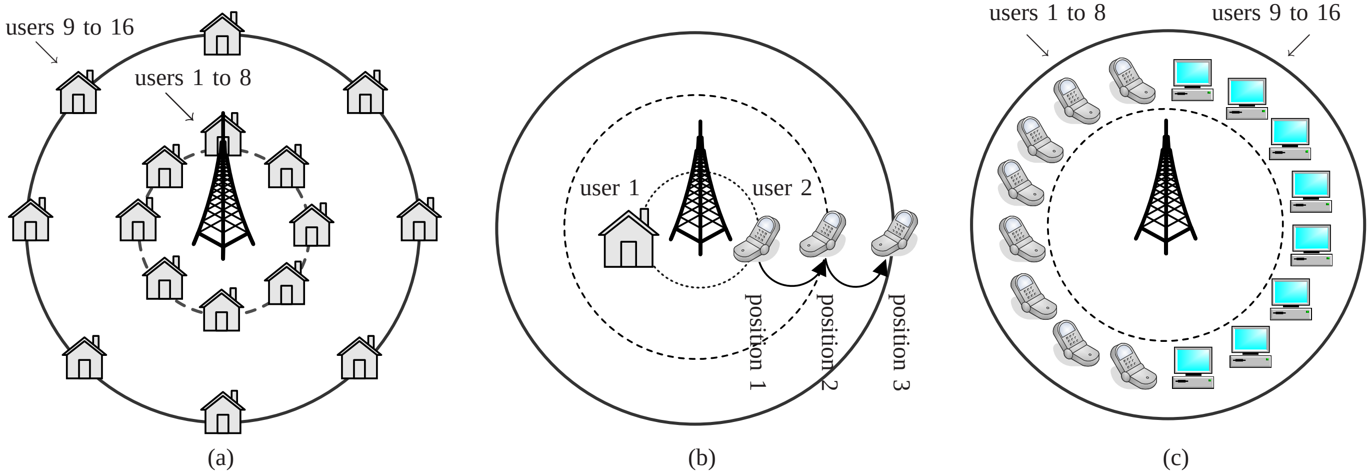


Fig. 3. Simulated scenarios: (a) fixed users, (b) a fixed user and a mobile user, (c) users with heterogeneous traffic

respectively.

$$Util_i\left(r_i\right) = \begin{cases} 0 & r_i \leq l_1, \\ sin^k\left(\frac{\pi}{2}\frac{r_i - l_1}{l_2 - l_1}\right) & l_1 < r_i \leq l_2, \\ 1 & r_i > l_2. \end{cases} \tag{43}$$

The simulated network consists of the BS, with total power equals to $20$ Watt, located at the center of the cell with 800m radius, that transmits accumulated traffic in its queues to the users over $64$ sub-carriers. The value of $T_c = 1000$ is chosen in this work, which is equivalent to $100$ frames or $1000$ symbols in IEEE 802.16[2] standard (downlink length= $1msec$, symbol length= $80\mu sec$ for $5MHz$ channel) [23]. We show the scheduling scheme performance for diverse channel status and traffic rates by considering the three scenarios shown in Fig. 3. In the first and the second scenarios, Fig. 3-(a) and Fig. 3-(b), the traffic is homogeneous, and we show the effect of channel gain variations on the scheduling performance. In the third scenario, Fig. 3-(c), we show the scheduling performance when users have heterogeneous traffic, i.e., users with non-concave and a concave utility functions exist.

[2]known as Worldwide Interoperability for Microwave Access (WiMAX)

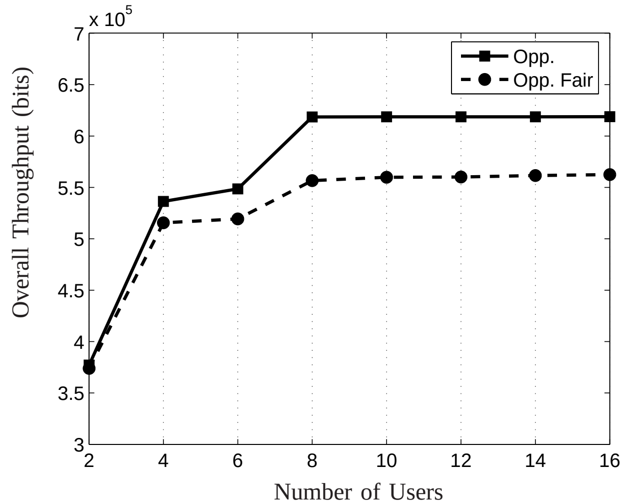


Fig. 4. Overall network throughput of the first scenario

## A. Fixed Users

In the first scenario, there are $16$ users, half of them are uniformly located on a circle with $50$ meters radius, and the other half are located on the cell edge at equal angular distance. Users have diverse channel gains due to path loss and distance fading. We investigate the effect of multi-user diversity on throughput and fairness performance of the scheduling schemes using this scenario.

Fig. 4 shows overall throughput of the network versus the number of users for the opportunistic and opportunistic fair scheduling schemes. As the opportunistic scheduling assigns a sub-carrier to a user with the highest channel gain, its throughput is the upper bound. The opportunistic fair scheduling achieves lower throughput than opportunistic scheduling because in some scheduling intervals it assigns a number of sub-carriers to users who have not been supported for a long time, irrespective to their channel gain. Both scheduling schemes exploit multi-user diversity as more users join the inner circle, i.e., when the number of users increases from $1$ to $8$ in Fig. 4. Users $9$ to $16$ are far from the BS and their channel gains are always much lower than the users located on inner circle, so they do not increase multi-user diversity gain and the throughput remains almost constant when these users join the network.

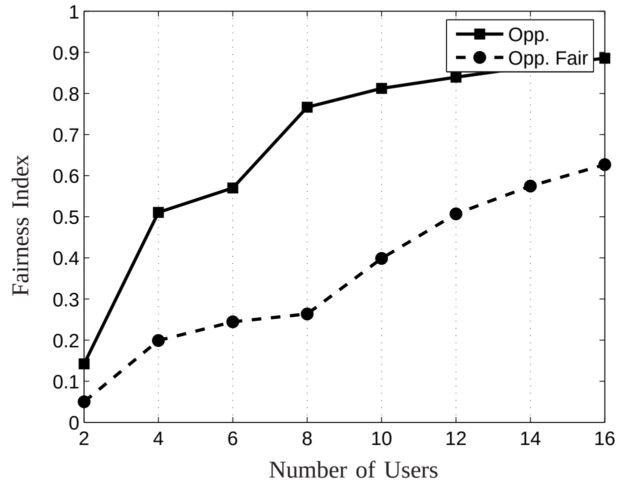


Fig. 5. Fairness index of the first scenario

Fig. 5 shows the Gini fairness index of the first scenario. The fairness index of opportunistic and opportunistic fair scheduling increases as the number of users increases. Increasing user diversity has an adverse effect on fairness. However, this effect is moderated in the opportunistic fair scheduling especially at low spatial diversity, i.e., users 1 to 8.

### *B. A Fixed and a Mobile User*

In the second scenario, a fixed user and a mobile user that moves away from the BS are considered. At first, users 1 and 2 are located close to the BS with the same distance. Then, user 2 moves away from the BS toward the edge of the cell. We investigate the adaptivity of the opportunistic fair scheduling in capturing the network status variations using this scenario.

Fig. 6 shows the throughput of user $1$ and user $2$ at three positions for opportunistic and opportunistic fair scheduling schemes. The throughput of opportunistic fair scheduling has been shown for two different time constants, $T_c$, of the exponentially weighted moving average. As user $2$ moves away from the BS and its average channel gain drops, the opportunistic scheduling allocates less rate to it and finally ignores it when it is very far. On the other hand, the opportunistic fair scheduling scheme, which intends to allocate proportional rates to the fair weights, allocates more rate to user $2$ than the ones of opportunistic allocation. In comparison to

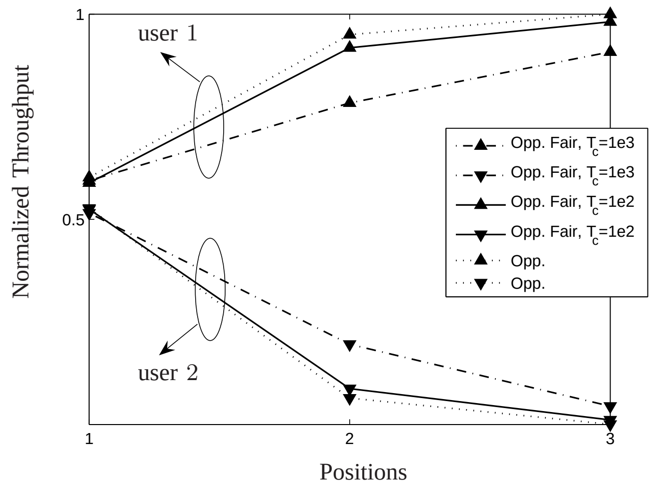


Fig. 6. User 1 and User 2's throughput at different positions of the second scenario

scheduling schemes with large $T_c$, the opportunistic fair scheduling scheme with small $T_c$ is less effective in compensating the effect of bad channel gain of user $2$ as it moves away from the BS. This can be explained as follows. A smaller number of scheduling intervals is considered and compensated for in the fairness scheme when $T_c$ is small. Therefore, the scheduler has shorter time to compensate for the unfairness.

Fig. 7 shows the Gini fairness index of the opportunistic and opportunistic fair scheduling with two different $T_c$ in the second scenario. When both users are close to the BS and their channels are almost similar, unfairness of opportunistic scheduling is not observed. However, as user $2$ moves and its channel condition degrades, the opportunistic fair scheduling treats it more fairly than the opportunistic scheduling, so the fairness index of the opportunistic scheduling deteriorates when user $2$ is at positions $2$ and $3$. Opportunistic fair scheduling with larger $T_c$ outperforms the one with smaller $T_c$ in terms of fairness.

The performance study of the second scenario indicates that the opportunistic fair scheduling can capture the network changes and adapt the fairness scheme accordingly. The adaptivity of the scheme can be adjusted by controlling the transmission history duration which is one of the components of the fairness module. Furthermore, the trade off between fairness and throughput

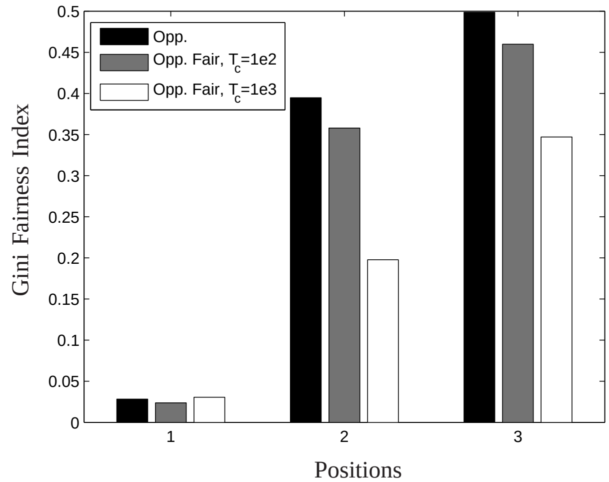


Fig. 7. Fairness performance of the second scenario

can be adjusted similarly.

### C. Heterogeneous Users

In the third scenario, all $16$ users are within the same distance from the BS, on a circle with $50$ meters radius, but they run two different applications with different utility functions. The first group of users, users $1$ to $8$, have a non-concave utility function, and the second group of users, users $9$ to $16$, have a concave utility function.

The utilities of users $1$ to $8$ versus time, when their traffic is scheduled by opportunistic and opportunistic fair scheduling schemes, are represented in Fig. 8-a and Fig. 8-b, respectively. The figures show that, first, opportunistic scheduling ignores few users with low channel gains over the simulation intervals, such as user $8$ in Fig. 8-a. This fact causes severe unfairness in service provisioning when user diversity is high. Second, the rate allocations and hence the users' utilities for opportunistic scheduling is highly interrupted in time compared to those of opportunistic fair scheduling. Although scheduling elastic traffic, with concave utility, is not sensitive to service provisioning delay, inelastic traffic, with non-concave utility, should be scheduled within least possible of service delays. Therefore, opportunistic scheduling is not appropriate for inelastic traffic service provisioning.

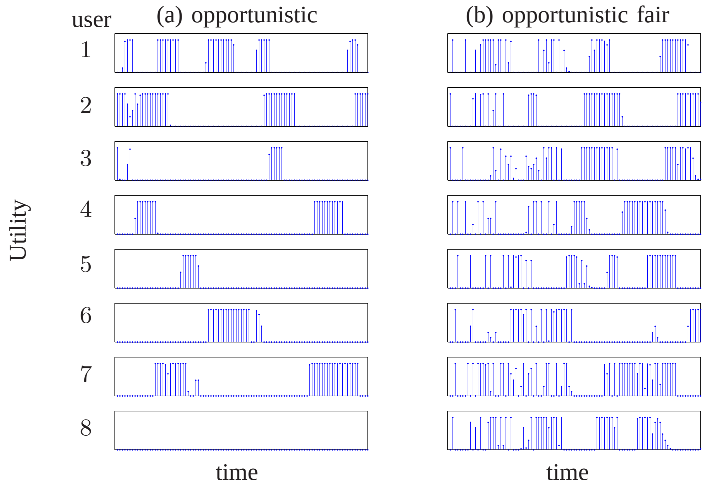


Fig. 8. Utilities of users 1 to 8 versus time for opportunistic and opportunistic fair scheduling schemes

Table II
AGGREGATE UTILITIES OF THE SCHEDULING SCHEMES

| **Scheduling Scheme** | $\sum_{i=1}^{8} U_i(r_i)$ | $\sum_{i=9}^{16} U_i(r_i)$ | $\sum_{i=1}^{16} U_i(r_i)$ |
|---|---|---|---|
| **Opportunistic** | 161.4702 | 141.0092 | 302.4793 |
| **Opportunistic Fair** | 306.2989 | 196.6107 | 502.9096 |

Furthermore, the aggregate users' utilities shown in Table II for both scheduling schemes demonstrate that the improvement in resource utilization or in the users' satisfaction of received service, represented by sum of the users' utilities, is higher for opportunistic fair scheduling than that of the opportunistic scheduling scheme. Moreover, the aggregate utilities of users with non-concave utilities are higher than that of the users with concave utilities. The reason is that the gradient of the non-concave utility function is higher than the gradient of the concave utility function at lower rates. Therefore, for the same allocated rate, the non-concave utility is larger than the concave utility.

## V. CONCLUSIONS

Fair weights have been designed for scheduling heterogeneous traffic in the downlink of OFDMA networks. We adopt the utility proportional fair criteria, design a set of fair weights associated with users, and propose an opportunistic fair scheduling which allocates the resources according to the fair weights. The proposed scheduler is adaptive because the fair weights can be modified dynamically when the network characteristics change due to mobility of users, admitting a new user, or changing the fairness policy of the network service provider. In addition, it reduces service interruption for real-time traffic which is sensitive to long service delays. In our further works, we will investigate various optimal strategies to detect changes in users' average channel gains which trigger the computation of the fair weights.

## VI. APPENDIX

The mathematical representations of $\nabla^2_{\mathbf{ww}}\mathcal{L}$ and $\nabla_{\mathbf{w}} f(\mathbf{w})$, required by the interior point algorithm and depended on users' utility functions, are presented in the appendix.

The objective function of $P_4$, based on utility functions (43), is given by:

$$f(\mathbf{w}) = \; - \; log(Util_1(w_1)) - \ldots - log(Util_M(w_M)). \tag{44}$$

Accordingly, $\nabla_{\mathbf{w}} f(\mathbf{w}) = \left(\frac{\partial f}{\partial w_{11}}, \ldots, \frac{\partial f}{\partial w_{MK}}\right)^T$ is computed as follows:

$$\nabla_{\mathbf{w}} f(\mathbf{w}) = - \begin{pmatrix} \frac{1}{Util_1(w_1)} \frac{\partial Util_1(w_1)}{\partial w_{11}} \\ \vdots \\ \frac{1}{Util_1(w_1)} \frac{\partial Util_1(w_1)}{\partial w_{MK}} \\ \vdots \\ \frac{1}{Util_M(w_M)} \frac{\partial Util_M(w_M)}{\partial w_{M1}} \\ \vdots \\ \frac{1}{Util_M(w_M)} \frac{\partial Util_M(w_M)}{\partial w_{MK}} \end{pmatrix}, \tag{45}$$

where, for $j = 1, \cdots, N_c$, and $\theta = \frac{\pi}{2} \frac{w_i - l_1}{l_2 - l_1}$:

$$\frac{\partial Util_{\breve{i}}}{\partial w_{ij}} = \begin{cases} \frac{-k\pi}{2(l_2 - l_1)} \frac{cos(\theta)}{sin(\theta))}, & if \quad i = \breve{i}, \\ 0, & otherwise. \end{cases} \tag{46}$$

To obtain $\nabla^2_{\mathbf{ww}}\mathcal{L}$, first $\nabla^2_{\mathbf{ww}}f(\mathbf{w})$ and $\nabla^2_{\mathbf{ww}}\mathbf{c}(\mathbf{w})$ are computed:

$$\nabla^2_{\mathbf{ww}}f(\mathbf{w}) = \; - \begin{pmatrix} \mathbf{G}(w_1) & 0_{(N_c,N_c)} & \dots & 0_{(N_c,N_c)} \\ 0_{(N_c,N_c)} & \mathbf{G}(w_2) & \dots & 0_{(N_c,N_c)} \\ \vdots & \vdots & & \vdots \\ 0_{(N_c,N_c)} & 0_{(N_c,N_c)} & \dots & \mathbf{G}_(w_M) \end{pmatrix}, \tag{47}$$

where

$$\mathbf{G}(w_i) = \begin{pmatrix} \frac{\partial^2 Util_i}{\partial w_{i1}\partial w_{i1}} & \dots & \frac{\partial^2 Util_i}{\partial w_{i1}\partial w_{iK}} \\ \frac{\partial^2 Util_i}{\partial w_{i2}\partial w_{i1}} & \dots & \frac{\partial^2 Util_i}{\partial w_{i2}\partial w_{iK}} \\ \vdots & & \vdots \\ \frac{\partial^2 Util_i}{\partial w_{iK}\partial w_{i1}} & \dots & \frac{\partial^2 Util_i}{\partial w_{iK}\partial w_{iK}} \end{pmatrix}. \tag{48}$$

$0_{(N_c,N_c)}$ is a $N_c \times N_c$ matrix with all zero entries. The second partial derivatives of the utility functions required for calculating $G(w_i)$ functions are:

$$\frac{\partial^2 Util_i}{\partial w_{\breve{i}\breve{j}}\partial w_{ij}} = \begin{cases} \frac{k\pi^2}{4(l_2-l_1)^2}\left(\frac{1}{sin^2(\theta)}\right), & if \quad i=\breve{i}, \\ 0, & otherwise, \end{cases} \tag{49}$$

for $\breve{j}$ and $j \in \{1, \cdots, N_c\}$.

In problem $P_4$, $\mathbf{c}(\mathbf{w})$ is represented by:

$$\mathbf{c}(\mathbf{w}) = \begin{pmatrix} \sum_{j=1}^{N_c} w_{1j} - \frac{l_1+l_2}{2} \\ \vdots \\ \sum_{j=1}^{N_c} w_{Mj} - \frac{l_1+l_2}{2} \\ P_{BS} - \sum_{j=1}^{N_c}\sum_{i=1}^{N_u} \frac{2^{w_{ij}}-1}{\alpha_{ij}} \\ w_{11} \\ \vdots \\ w_{MK} \end{pmatrix}. \tag{50}$$

Accordingly, $\nabla^2_{\mathbf{ww}}\mathbf{c}(\mathbf{w})$ for calculating $\nabla^2_{\mathbf{ww}}\mathcal{L}$ can be obtained by:

$$\nabla^2_{\mathbf{ww}}\mathbf{c}(\mathbf{w}) = (ln(2))^2 \begin{pmatrix} \frac{2^{w_{11}}}{\alpha_{11}} & 0 & \dots & 0 \\ 0 & \frac{2^{w_{12}}}{\alpha_{12}} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \frac{2^{w_{MK}}}{\alpha_{MK}} \end{pmatrix}. \tag{51}$$

## References


[1] R. Knopp and P. A. Humblet, "Information Capacity and Power Control in Single Cell Multiuser Communications," in *Proc. IEEE International Conference on Communications (ICC'95), Seattle, USA*, June 1995.

[2] "A framework for opportunistic scheduling in wireless networks," *Comput. Netw.*, vol. 41, no. 4, pp. 451 – 474, 2003.

[3] Y. Liu and E. Knightly, "Opportunistic Fair Scheduling over Multiple Wireless Channels," in *Proc. IEEE International Conference on Computer Communications (INFOCOM'03), San Francisco, USA*, April 2003.

[4] Z. Zhang, Y. He, and E. K. P. Chong, "Opportunistic Scheduling for OFDM Systems with Fairness Constraints," *EURASIP J. Wirel. Commun. Netw.*, vol. 8, no. 3, pp. 275–277, 2008.

[5] A. Wang, L. Xiao, S. Zhou, X. Xu, and Y. Yao, "Dynamic resource management in the fourth generation wireless systems," in *Proc. International Conference on Communication Technology (ICCT'03), Beijing, China*, April 2003.

[6] A. Jalali and R. Padovani, "Data Throughput of CDMA-HDR a High Efficiency-High Data Rate Personal Communication Wireless System," in *Proc. IEEE Vehicular Technology Conference (VTC 2000-Spring), Tokyo, Japan*, Spring 2000.

[7] I. C. Wong and B. L. Evans, "Optimal Downlink OFDMA Resource Allocation with Linear Complexity to Maximize Ergodic Rates," *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 962–971, 2008.

[8] H. T. Cheng and W. Zhuang, "An Optimization Framework for Balancing Throughput and Fairness in Wireless Networks with QoS Support," in *Proc. the 3rd International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QShine'06), Waterloo, Ontario, Canada*, 2006.

[9] M. Awad, V. Mahinthan, M. Mehrjoo, X. Shen, and J. Mark, "Downlink Resource Allocation for OFDMA-based Multiservice Networks with Imperfect CSI," in *Proc. IEEE International Conference on Communications (ICC'09), Dresden, Germany*, June 2009.

[10] F. Kelly, A. Maulloo, and D. Tan, "Rate Control in Communication Networks: Shadow Prices, Proportional Fairness and Stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, 1998.

[11] J. Mo and J. Walrand, "Fair End-to-End Window-Based Congestion Control," *IEEE/ACM Trans. Networking*, vol. 8, no. 5, pp. 556–567, 2000.

[12] M. Dianati, X. Shen, and K. Naik, "Coperative Fair Scheduling for the Downlink of CDMA Cellular Networks," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 1749–1760, 2007.

[13] T. Zhang, Z. Zeng, and Y. Qiu, "A subcarrier allocation algorithm for utility proportional fairness in ofdm systems," may 2008, pp. 1901 –1905.

[14] M. Mehrjoo, M. Dianati, X. Shen, and K. Naik, “Opportunistic Fair Scheduling for the Downlink of IEEE 802.16 Wireless Metropolitan Area Networks,” in *Proc. the 3rd International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QShine), Waterloo, Canada*, August 2006.

[15] M. Mehrjoo, S. Moazeni, and X. Shen, “A New Modeling Approach for Utility-Based Resource Allocation in OFDM Networks,” in *Proc. IEEE International Conference on Communications (ICC’08), Beijing, China*, May 2008.

[16] K. Seong, M. Mohseni, and J. M. Cioffi, “Optimal Resource Allocation for OFDMA Downlink Systems,” in *Proc. IEEE International Symposium on Information Theory (ISIT’06), Seattle, USA*, July 2006.

[17] W. Yu and R. Lui, “Dual Methods for Nonconvex Spectrum Optimization of Multicarrier Systems,” *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, 2006.

[18] J. Nocedal and S. Wright, *Numerical Optimization*. Springer, 2006.

[19] M. Mehrjoo, M. K. Awad, and X. Shen, “Resource Allocation for OFDM-Based WiMAX,” in *WiMAX network planning and optimization*. CRC Press - Taylor and Francis, 2009.

[20] J. W. Mark and W. Zhuang, *Wireless Communications and Networking*. Upper Saddle River, New Jersey, USA: Pearson Education International, 2003.

[21] J. Cai, X. Shen, and J. Mark, “Downlink Resource Management for Packet Transmission in OFDM Wireless Communication Systems,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1688–1703, 2005.

[22] D. M. Chiu and A. S. Tam, “Fairness of Traffic Controls for Inelastic Flows in the Internet,” *Comput. Netw.*, vol. 51, no. 11, pp. 2938–2957, 2007.

[23] IEEE, *IEEE 802.16e-2005*, IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems - Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands, Std., February 2006.