High-Throughput and Fair Scheduling for Access Point Cooperation in Dense Wireless Networks

Mengyao Ge and *Douglas M. Blough*
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332-0250
{mengyao, doug.blough}@ece.gatech.edu

***Abstract*—This paper studies the fair scheduling problem for dense wireless networks with AP cooperation and MIMO links. The problem is to maximize the aggregate throughput while meeting a specified fairness objective. The proposed scheduling algorithm works with a novel single-slot throughput optimization procedure to first generate a set of candidate communication sets that are both high-performing and have good representation among all users. For a given set of candidate communication sets, our algorithm then produces a communication schedule that achieves near-optimal aggregate throughput among solutions that meet the specified fairness objective. Simulation results show that our proposed scheduling algorithm achieves high aggregate throughput while maintaining very good fairness. We also include a mechanism to relax the fairness constraint by a specified amount, which produces a range of solutions that permit a trade-off between performance and fairness.**

I. Introduction

In current dense wireless networks, co-channel interference is a major impediment that prevents users from achieving the full benefits of Multiple-Input Multiple-Output (MIMO) transmission capacity. This situation is further exacerbated by the proliferation of both access points (APs) and devices, as well as by unplanned deployments. Traditional techniques, such as assigning orthogonal channels to different APs and assigning non-overlapping time slots to different users for 802.11-based WLAN, at best equally divide available bandwidth among users. A potential way to break the bottleneck of performance in dense wireless networks is AP cooperation combined with advanced MIMO processing techniques [1].

A common scenario is that a large number of APs are part of one enterprise wireless network, which provides strong opportunities for AP cooperation. We mainly focus on indoor wireless communications, where most users are expected to be stationary for significant periods of time. This scenario is consistent with most enterprise environments. Since there are many users sharing the limited resources of the wireless network, link scheduling arises as a key problem, i.e. determining how to activate links for a given scheduling period to meet the desired organizational requirements. The maximum-performance scheduling problem with fairness constraints in dense MIMO networks is the subject of this paper.

In general, high throughput and optimal fairness are two fundamental objectives in wireless networks that cannot be maximized simultaneously. This motivates the investigation of inherent tradeoffs between the two objectives, where a common approach is to maximize performance subject to some fairness constraints. To avoid the ”performance anomaly” with rate-based fairness [2], time-based fairness is widely used in multi-rate wireless networks. The basic idea is to allocate equal time to each user and the bandwidth of each user is then dependent on the number of users and its own data rate [3].

In this paper, we address the fair scheduling problem within a small group of APs that are operating in the same wireless channel. APs are coordinated such that they jointly process the user data signals. Our approach is to let APs share physical layer information and schedule a jointly-optimized set of communications at each time slot, which can ultimately meet the fairness objective over an entire round of communications. Our main contributions are as follows:

- **Generating a diverse set of high-performing communication sets:** We utilize a recently-developed algorithm that jointly determines the stream allocation and precoding and combining weights to maximize weighted sum rate. User selection is built into this PHY-layer algorithm. This algorithm is used in conjunction with a procedure to adapt user weights over multiple iterations in order to produce a set of candidate communication sets that are both high performing and have good representation among the users.
- **Developing a high-performing time-based fair schedule:** We formulate a fair scheduling problem that takes as input a set of candidate communication sets and rates achieved for each user within a given communication set (the outputs of our communication set generation algorithm). The problem is to maximize aggregate rate subject to a very general fairness constraint, namely a set of desired bandwidths for each user, which can be used to incorporate a wide variety of fairness constraints into the problem. We are able to produce a near-optimal and efficient solution to this problem based on a linear-programming relaxation of the specified integer linear programming problem. We demonstrate the high performance and fairness achieved by the algorithm using a fairness constraint that is a generalization of the notion of interference-aware time-based fairness [4].
- **Throughput-fairness tradeoff:** To accommodate various

network requirements, we formulate the problem as an ϵ-approximate fair scheduling problem. The ϵ factor allows control of how tightly the solution achieves the specified fairness objective. This produces a range of solutions that allow for controlled tradeoffs between throughput and fairness. This would be very useful in scenarios where performance is critical and fairness goals are malleable.

II. Scheduling Framework

We consider a scenario in which single-hop wireless networks are densely deployed over a region, where the areas served by different access points (APs) can overlap. We expect that most users are stationary for significant periods of time and assign the time slots for those users with well-characterized channels. This is a common scenario for enterprise WLAN settings, which typically cover office-type environments. The durations of stationary periods are assumed to be much longer than the scheduling period, which is tens of milliseconds or less for the scenario considered herein. Although we do not develop solutions to explicitly deal with mobile users in the network, we do evaluate, through simulations, the impact of a minority of mobile users on the performance of our solutions. We focus on downlink transmissions since in typical indoor environments 80% or more of the traffic is on the downlink. Scheduling downlink or uplink traffic only within single time slots helps reduce channel estimation overhead as shown in [5].

A. AP cooperation

To address the interference problem in overlapping single-hop wireless networks, we consider the use of advanced MIMO techniques involving AP cooperation and coordination of communications across cells, which are envisioned to be widely used in next-generation wireless technologies. The complexity of coordination, backhaul limitations, and computational limits for scheduling will impose relatively small upper bounds on the number of APs that cooperate. Hence, we assume a large enterprise wireless network is clustered into small groups of APs that are near each other and operate on the same channel.[1] We assume that there is a single entity for each cluster, which has access to channel state information (CSI) and the data signals intended for all users, that computes the overall schedule and the precoding and combining weights for all APs and users active within each slot. This entity could be one designated AP or a network controller connected to all APs within a cluster.

B. Physical-layer model with linear transceiver design

Through the co-processing by multiple APs in the same cluster, interference within the cluster is eliminated. The interference mitigation is mainly implemented by the precoder and combiner design.

Assume there are M APs in one cluster with N_t antennas for each and K users with N_r antennas for each. The matrix of complex channel gains between the APs and the k^{th} user is denoted by $\boldsymbol{H}_k \in \mathbb{C}^{N_r \times MN_t}$. The data vector $\boldsymbol{x} = \left[\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_K^T\right]^T$ is assumed to be independently encoded Gaussian codebook symbols with $\mathbb{E}[\boldsymbol{x}_k \boldsymbol{x}_k^\dagger] = \boldsymbol{I}$, where $(\cdot)^\dagger$ is the conjugate transpose of $(\cdot)$. $\boldsymbol{V}_k \in \mathbb{C}^{MN_t \times N_r}$ and $\boldsymbol{U}_k \in \mathbb{C}^{N_r \times N_r}$ are the precoder for the signals of the k^{th} user.

The instantaneous data rate of the k^{th} user is given by

$$R_k = \log \left| \boldsymbol{I} + (\boldsymbol{U}_k^\dagger \boldsymbol{R}_{\bar{k}}^{-1} \boldsymbol{U}_k)(\boldsymbol{U}_k^\dagger \boldsymbol{H}_k \boldsymbol{V}_k \boldsymbol{V}_k^\dagger \boldsymbol{H}_k^\dagger \boldsymbol{U}_k) \right| \qquad (1)$$

where $\boldsymbol{R}_{\bar{k}}$ is the covariance matrix of the received interference plus noise.

C. Link scheduling approach

Since the channels in our target scenarios are assumed to be fixed for a moderate amount of time, we can assign a jointly-optimized set of communications for each time slot to obtain high aggregate performance, while achieving fairness among the competing users. Our basic approach is to have the scheduler generate a set of diverse and high-performance communication sets with calculated MIMO weights and data rates, after collecting the CSI and user data from all APs. Next, the scheduler computes a communication schedule that specifies the number of time slots allocated for each communication set in order to achieve a given fairness objective.

III. Communication Sets Generation with Linear MIMO Processing

In this section, we present the algorithm to generate a diverse set of high-performing communication sets. First, an iterative algorithm is proposed to solve the weighted sum-rate maximization (WSRM) problem, which can jointly determine the subset of users to be activated, the number of streams transmitted to each user and the MIMO weights. Then, we adjust the user weights and perform the iterative algorithm for a set of WSRM problems to generate a number of communication sets with the achieved rates. The WSRM problem with per-node power constraint is formulated as follows,

$$\begin{array}{ll} \max & \sum_{k=1}^{K} w_k R_k(\boldsymbol{U}, \boldsymbol{V}) \\ s.t. \;\; C_0: & Tr(\boldsymbol{S}_m \boldsymbol{V} \boldsymbol{V}^\dagger) \leq P_m, m = 1, \ldots, M \end{array} \qquad (2)$$

where $w_k \geq 0$ is the weight assigned to the k^{th} user's rate and P_m is the maximum transmit power of the m^{th} AP. A diagonal matrix $\boldsymbol{S}_m \in \mathbb{R}^{N_t \times N_t}$ is introduced for each AP, in order to select the partition of precoding matrix $\boldsymbol{V}$ applied at the m^{th} AP. Thus, $\boldsymbol{S}_m$ carries ones on the diagonal elements corresponding to the antennas of the m^{th} AP and zeros otherwise.

For a given set of linear precoder $\boldsymbol{V}_k$'s, the optimal combiner applied at the receiver k is given by the MMSE combiner [6],

$$\boldsymbol{U}_k = (\boldsymbol{H}_k \boldsymbol{V}_k \boldsymbol{V}_k^\dagger \boldsymbol{H}_k^\dagger + \boldsymbol{R}_{\bar{k}})^{-1} \boldsymbol{H}_k \boldsymbol{V}_k, \qquad (3)$$

which is information lossless.

[1] Our techniques can be applied independently across as many orthogonal channels as are available in a given wireless deployment.

TABLE I: Algorithm for communication sets generation

	Iterative Algorithm for solving WSRM
1:	Initialization: $\boldsymbol{V}_k = \boldsymbol{V}_k^0$ for all $k \in [1, \dots, K]$;
2:	Repeat
3:	Compute $\boldsymbol{U}_k$ using (3) for all $k \in [1, \dots, K]$ for given $\boldsymbol{V}_k$'s;
4:	Compute $\boldsymbol{W}_k$ using (5) for all $k \in [1, \dots, K]$ for given $\boldsymbol{V}_k$'s and $\boldsymbol{U}_k$'s;
5:	Update $\boldsymbol{V}_k$'s by solving problem (4);
6:	until $\left\lvert \sum_{k=1}^{K} \omega_k R_k^{(n)} - \sum_{k=1}^{K} \omega_k R_k^{(n-1)} \right\rvert \le \varepsilon_1$.

A. Solving the WSRM problem

To solve the WSRM problem, we have proposed, in a companion submission [7], an iterative algorithm using techniques similar to [8] [9], as summarized in Table I. In order to make this paper self-contained, we provide a very condensed description of this algorithm here.

The key idea of the algorithm is to solve a set of weighted sum mean-square-error (MSE) minimization problem iteratively to obtain a local WSR-optimum. The weighted MSE minimization problem is formulated as follows:

$$\max \quad \sum_{k=1}^{K} Tr(\boldsymbol{W}_k \boldsymbol{E}_k) \quad s.t. \;\; C_0, \tag{4}$$

where $\boldsymbol{E}_k$ is the MSE covariance matrix of the k^{th} user. It can be proved that the gradient of the WSR and weighted sum MSE are equal with MMSE combiner at the user side if

$$\boldsymbol{W}_k = w_k(\boldsymbol{I} + \boldsymbol{\Lambda}_k), \tag{5}$$

where $\boldsymbol{\Lambda}_k \in \mathbb{R}^{N_r \times N_r}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{V}_k^\dagger \boldsymbol{H}_k^\dagger \boldsymbol{R}_{\bar{k}}^{-1} \boldsymbol{H}_k \boldsymbol{V}_k$ ordered in decreasing order.

Based on these facts, an iterative algorithm that alternately calculates the combiner and precoder can be utilized to find a local WSR-optimum. The algorithm is given in Table I.

However, the solutions to the weighted sum MSE minimization problem in [8] [9] lack the ability to completely deactivate links efficiently. In other words, these solutions are unable to determine a certain subset of users as a candidate communication set. Therefore, we propose a different algorithm to solve problem (4), which jointly optimizes the precoding matrices and the number of active streams for each user. The candidate communication set after a given optimization execution is the set of users that have one or more active streams.

In the first step of our algorithm, the Lagrangian of the weighted sum MSE minimization problem is given by

$$L = \sum_{k=1}^{K} Tr(\boldsymbol{W}_k \boldsymbol{E}_k) + \sum_{m=1}^{M} \mu_m (Tr(\boldsymbol{S}_m \boldsymbol{V}\boldsymbol{V}^\dagger) - P_m), \tag{6}$$

where $\mu_m \ge 0, m = 1, \dots, M$ are the Lagrange multipliers.

Based on the KKT conditions, the gradient of L with respect to $\boldsymbol{V}_k^\dagger$ should be zero, which yields the following equation

$$\begin{aligned} \boldsymbol{H}_k^\dagger \boldsymbol{U}_k \boldsymbol{W}_k &= \boldsymbol{H}_k^\dagger \boldsymbol{U}_k \boldsymbol{W}_k \boldsymbol{U}_k^\dagger \boldsymbol{H}_k \boldsymbol{V}_k \\ &+ \sum_{l=1, l\neq k}^{K} \boldsymbol{H}_l^\dagger \boldsymbol{U}_l \boldsymbol{W}_l \boldsymbol{U}_l^\dagger \boldsymbol{H}_l \boldsymbol{V}_k + \sum_{m=1}^{M} \mu_m \boldsymbol{S}_m \boldsymbol{V}_k. \end{aligned} \tag{7}$$

To solve the precoder in (3) and (7), let us first perform the following compact SVD,

$$\boldsymbol{R}_{\bar{k}}^{-1/2} \boldsymbol{H}_k \boldsymbol{\Pi}_{\bar{k}}^{-1/2} = \boldsymbol{F}_k \boldsymbol{D}_k \boldsymbol{G}_k^\dagger \tag{8}$$

where

$$\boldsymbol{\Pi}_{\bar{k}} = \sum_{l=1, l\neq k}^{K} \boldsymbol{H}_l^\dagger \boldsymbol{U}_l \boldsymbol{W}_l \boldsymbol{U}_l^\dagger \boldsymbol{H}_l + \sum_{m=1}^{M} \mu_m \boldsymbol{S}_m; \tag{9}$$

and $\boldsymbol{D}_k \in \mathbb{R}^{N_r \times N_r}$ is a diagonal matrix containing the singular values of $\boldsymbol{R}_{\bar{k}}^{-1/2} \boldsymbol{H}_k \boldsymbol{\Pi}_{\bar{k}}^{-1/2}$ ordered in decreasing order; $\boldsymbol{F}_k \in \mathbb{C}^{N_r \times N_r}$ and $\boldsymbol{G}_k \in \mathbb{C}^{N_t \times N_r}$ are the corresponding left and right singular vectors of $\boldsymbol{R}_{\bar{k}}^{-1/2} \boldsymbol{H}_k \boldsymbol{\Pi}_{\bar{k}}^{-1/2}$.

Based on [10] [11], the precoding matrices to solve (3) and (7) have the following structure without loss of generality, For given μ_m's, the precoding matrix for receiver k that solves (3) and (7) is given by

$$\boldsymbol{V}_k = \boldsymbol{\Pi}_{\bar{k}}^{-1/2} \boldsymbol{G}_k \boldsymbol{\Psi}_k \tag{10}$$

where $\boldsymbol{\Psi}_k$ is a $N_r \times N_r$ diagonal matrix with nonnegative elements on the diagonal given by

$$\boldsymbol{\Psi}_k = \left(\boldsymbol{W}_k^{1/2} \boldsymbol{D}_k^{-1} - \boldsymbol{D}_k^{-2} \right)_+^{1/2}, \tag{11}$$

and $(\cdot)_+$ is the matrix $(\cdot)$ with the negative elements replaced with zeros. Here, the $(\cdot)_+$ operation in component $\boldsymbol{\Psi}_k$ can potentially turn off some streams and thus achieve the stream allocation and user selection. The Lagrangian multiplier μ_m's can be updated using subgradient method or ellipsoid method.

B. Adjusting the link weights

Let $\boldsymbol{b}$ be a $K \times 1$ vector, where the k^{th} element b_k represents the desired bandwidth portion allocated to the k^{th} user. Different fairness objectives can be achieved by choosing the corresponding parameter $\boldsymbol{b}$. If we aim to generate N communication sets, let $\boldsymbol{r}_k$ be a $1 \times N$ vector that contains the data rates of the k^{th} user, i.e. $r_{k,n}$ denotes the data rate of the k^{th} user in the n^{th} communication set. When generating the $(i+1)^{\text{th}}$ communication set after the first i sets have already been generated, the basic idea is to assign larger weights to users that are below their desired bandwidth proportions when considering the first i sets. A user k that is at or above its desired bandwidth proportion is assigned weight $w_k = 0$ and is therefore excluded from the current communication set by design. This approach yields satisfying results in balancing high-performance communication sets and incorporating user diversity into the chosen high-performing sets. Mathematically, there are various ways to achieve the aforementioned weight adjustment idea. A general form is

$$w_k \begin{cases} > w_j & \text{if} \quad 0 \le u_k/b_k < u_j/b_j \le 1 \\ = 0 & \text{if} \quad u_k/b_k \ge 1 \end{cases} \tag{12}$$

where u_k is the bandwidth proportion of the kth user from previously computed N communication sets, given by $u_k = \frac{\sum_{n=1}^{N} r_{k,n}}{\sum_{k=1}^{K} \sum_{n=1}^{N} r_{k,n}}$. In this paper, we set the link weights as follows:

$$w_k = \max\left(1 - u_k/b_k, 0\right). \tag{13}$$

C. Single-user MIMO communication sets

To fully exploit the communication set diversity, we also compute communication sets with a single active user per set. In this case, the active user achieves its interference-free data rate and is jointly served by the cooperative APs. The interference-free data rate of the single user can be obtained by applying a SVD precoder with optimal power allocation.

IV. MIMO Link Scheduling

After generating $N_{tot} = N + K$ communication sets as discussed in Sec. III, our focus is on developing a proportionally fair schedule to provide both high aggregate performance and specified fairness. Let $\mathcal{G}_i$ be the i^{th} user set containing user $\{v_{1,i}, \ldots, v_{n_i,i}\}$ with data rate $\{r_{1,i}, \ldots, r_{n_i,i}\}$. In this section, we formulate a throughput-optimal scheduling problem with fairness constraint based on a set of N_{tot} candidate communication sets. We also show how this problem can be solved nearly optimally and efficiently with linear programming.

Let $x_i \geq 0$ be the number of time slots to schedule the i^{th} communication set $\mathcal{G}_i$. The data rate of user j over the schedule period T_s can be represented as

$$R_j = \sum_{i:j\in\mathcal{G}_i} r_{j,i} x_i / T_s$$

We wish to find a vector $\boldsymbol{x} \in \mathbb{Z}^{N\times 1}$ that satisfies the proportional fairness constraint

$$C_1 : R_1 : R_2 : \cdots : R_K = b_1 : b_2 : \cdots : b_K \ ,$$

For a given schedule duration, T_s, containing N_{slot} time slots, we have the following constraint:

$$C_2 : \sum_{i=1}^{N_{tot}} x_i \leq N_{slot}$$

The scheduling problem is then formulated to maximize the aggregate throughput under the fairness constraint,

$$\begin{aligned} \max_{\boldsymbol{x}} \quad & \sum_{j=1}^{K} R_j T_s \\ s.t. \quad & x_i \geq 0, x_i \in \mathbb{Z}, i = 1, \ldots, N_{tot} \\ & C_1 - C_2 \end{aligned} \tag{14}$$

Note that the fairness constraint C_1 contains $K-1$ equality constraints (i.e., $\frac{R_j}{b_j} = \frac{R_1}{b_1}$, for $j = 2, \ldots, K$). However, the perfect fairness imposed by C_1 lacks the flexibility to accommodate different scenarios. Therefore, we introduce the notion of ϵ-approximate fairness and relax the C_1 into a set of inequality constraints

$$C_3 : c(1-\epsilon)b_j \leq R_j \leq c(1+\epsilon)b_j, j = 1, \ldots, K. \tag{15}$$

where $c = \sum_{j=1}^{K} R_j$. By replacing C_1 in (14) with C_3, a new scheduling problem with a variable fairness objective is formulated. Note that if $\epsilon = 0$, C_3 becomes the same as C_1, which leads to perfect fairness. When ϵ becomes large, the scheduling problem corresponds to a throughput maximization problem with no fairness constraint.

TABLE II: Scheduling Algorithm

Interior-point method
input: data rates in candidate communication sets $\{r_{i,j}\}_{i:j\in\mathcal{G}_\rangle}$, desired bandwidth proportion $\boldsymbol{b}$
output: optimized schedule $\boldsymbol{s}^* = \{s_1^*, \ldots, s_N^*\}$
1. Initialization: Given feasible $\boldsymbol{x} = \boldsymbol{x}^0$, $t := t^{(0)} > 0$, $\mu > 0$, ε.
2. **Repeat**
3. Compute the optimal solution $\boldsymbol{x}^*(t)$ to (18) starting at $\boldsymbol{x}$.
4. Update $\boldsymbol{x} := \boldsymbol{x}^*(t)$.
5. Quit if $(2K + N - 1)/t < \varepsilon$.
6. Update $t := \mu t$.
7. $\boldsymbol{s}^* = \text{round}(\boldsymbol{x})$.

A. Scheduling algorithm

A general way to solve the formulated integer linear programming (ILP) problem (14) is to solve its LP relaxation and then round the entries of the solution. The LP relaxation of (14) is given by

$$\begin{aligned} \max_{\boldsymbol{x}} \quad & \sum_{j=1}^{K} \sum_{i:j\in\mathcal{G}_i} r_{j,i} x_i \\ s.t. \quad & x_i \geq 0, i = 1, \ldots, N_{tot} \\ & C_2 - C_3 \end{aligned} \tag{16}$$

Since the problem (16) is a standard LP problem, it can be addressed by well-known techniques such as the interior-point method. The relaxed LP has at least one feasible solution, due to the existence of single-user communication sets.

To solve the formulated scheduling problem, we introduce a simple interior-point method, called the barrier method.

First, we define a logarithmic barrier function $\phi(\boldsymbol{x})$

$$\begin{aligned} \phi(\boldsymbol{x}) \quad &= -\sum_{i=1}^{N_{tot}} \log(x_i) - \log(N_{slot} - \sum_{i=1}^{N_{tot}} x_i) \\ &- \sum_{j=1}^{K} \log\left(c(1+\epsilon)b_j - \sum_{i:j\in\mathcal{G}_i} r_{j,i} x_i / T_s \right) \\ &- \sum_{j=1}^{K} \log\left(\sum_{i:j\in\mathcal{G}_i} r_{j,i} x_i / T_s - c(1-\epsilon)b_j \right) \end{aligned} \tag{17}$$

Next, we define an unconstrained minimization problem with parameter t,

$$\min_{\boldsymbol{x}} f_t(x) = -t \sum_{j=1}^{K} \sum_{i:j\in\mathcal{G}_i} r_{j,i} x_i + \phi(\boldsymbol{x}) \tag{18}$$

The optimal solution to problem (18) is an approximation of the optimal solution to problem (16), where $t > 0$ is a parameter that sets the accuracy of the approximation. As t increases, the approximation becomes more accurate. The outline of barrier method is summarized in Table II. To solve the unconstrained minimization problem (18) in each iteration, Newton's method can be used to compute the optimal solution.

Once we have the optimal solution to the LP problem, we perform the following rounding procedure. First, sort the $res_i = x_i^* - \lfloor x_i^* \rfloor$ in descending order. Then, the solution x_i^*'s of the first I user sets with the higher res_i value will be rounded to $\lceil x_i^* \rceil$, while the remaining x_i^*'s will be rounded to $\lfloor x_i^* \rfloor$, where $I = N_{slot} - \lfloor x_i^* \rfloor$. The rounded solution determines the number of time slots assigned to each communication set. Although the fairness constraint C_3 might be violated after the rounding procedure, it will be demonstrated in the simulations that any deviation from the targeted fairness is quite small.

B. Achieving fairness

Different choices of parameter $\boldsymbol{b}$ achieve different fairness objectives, which can represent various QoS requirements. In this paper, we are particularly interested in achieving time-based fairness since it has been shown to be particularly well-suited for multi-rate wireless networks. In [4], the idea of time-based fairness is extended to take interference into account. Specifically, each user is allocated an equal number of interference-free time slots, where its bandwidth then depends on the number of users and its own channel quality. Different from the standard notion of time-based fairness in wireless networks, this fairness notion eliminates interference-induced distortions on data rates introduced by the scheduling algorithm. The desired bandwidth portion is defined as $b_k = \rho_k / \sum_{k=1}^{K} \rho_k, \forall k$, where ρ_k is the interference-free data rate as discussed in Sec. III-C.

Another well known fairness metric is called rate-based fairness, which aims to achieve equal average rates across all users. This criterion can be realized by setting $b_k = 1/K, \forall k$. However, rate-based fairness degrades the throughput of all users in order to match the throughput of the user with the lowest channel quality. Although it is known to produce lower overall performance than time-based fairness, we include rate-based fairness as a comparison point as well.

To evaluate the fairness, we use the fairness index proposed in [4],

$$FI(\boldsymbol{u}, \boldsymbol{b}) = \exp\left(-\frac{1}{K}\sum_{k=1}^{K}\left|\ln\frac{u_k}{b_k}\right|\right), \qquad (19)$$

where u_k is the fraction of bandwidth allocated to the i^{th} user. The fairness index given by (19) takes values in $[0, 1]$, with 1 representing perfect fairness among users.

V. Simulation experiments

In this section, we report on simulation experiments to evaluate the performance of our proposed scheduling algorithm, which we denote by **TimeFair** in this section. The optimal solution to the LP relaxation problem is referred to as **Relaxed TimeFair**, which serves as an upper bound of the ILP problem. For comparison, we also consider the following algorithms:

- **TDMA**: This is a basic time-fair TDMA scheduling algorithm, where the links are scheduled sequentially in a round robin manner. Since in each time slot, there is only one user scheduled and served by all APs, it can achieve the interference-free data rates using the SVD MIMO weights. Moreover, TDMA allocates the bandwidth with perfect fairness in a time-based sense.
- **GreedyMaxRate**: This is a scheduling algorithm that aims to maximize the throughput but with only minimal fairness. Here, the minimal fairness indicates that each user is allocated at least one time slot. This algorithm will be seen to achieve very high aggregate performance since it is relatively unconstrained by fairness concerns.
- **RateFair**: This algorithm works similar to our proposed time-based fair scheduling algorithm, the only difference being that all the users are assigned with the same desired bandwidth proportion, i.e. $b_k = 1/K$ for all k.

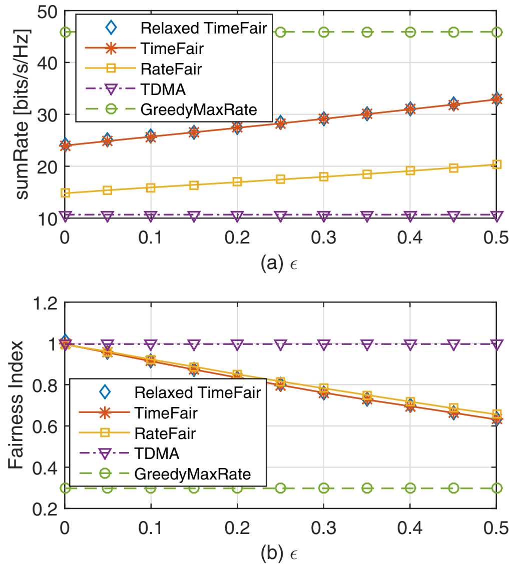

Fig. 1: Throughput and fairness as a function of ϵ.

A. Simulation setup

The simulation experiments are performed as follows. We consider that the access points (APs) are located in a line with an interval of 30 meters. There are k_m receivers uniformly distributed around the m^{th} AP within a radius of 50 meters with $\sum_{m=1}^{M} k_m = K$. We set each AP to have 4 antenna elements and each user to have 2 antenna elements. To compute the SNR and SINR values, we use a quasi-static Rayleigh flat-fading channel model with a path-loss exponent of 3 and the noise power of -80 dBm. The transmit power of each AP is 20 dBm. Throughout the following experiments, 3 coordinated APs are located in a line with 30 m interval. The scheduling period is set to 2 s and the duration of each time slot is 10 ms. All presented results are averaged over 500 random user deployments.

B. Simulation results

We first present the results obtained when there are 15 users per AP for a total of 45 users. Note that the number of supportable users can be multiplied by the number of available orthogonal channels. Fig. 1 shows the sum rate and fairness index achieved by different algorithms, where the fairness factor ϵ is varied from 0 to 0.5. As the fairness requirement is relaxed, the sum-rates of both TimeFair and RateFair increase in Fig. 1(a) while their fairness indices decrease in Fig. 1(b). The sum-rate gap of about 35% between TimeFair

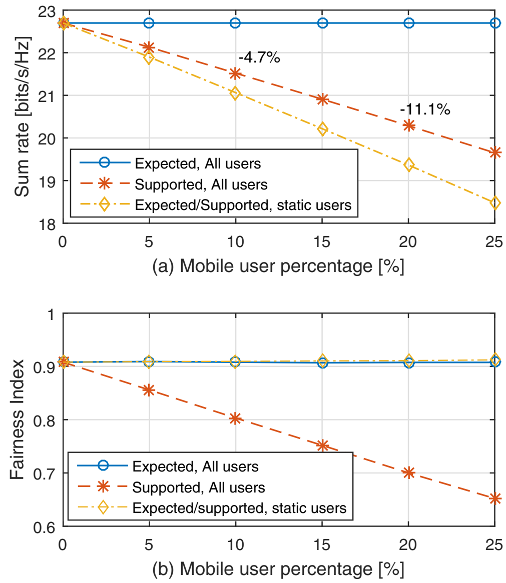

Fig. 2: Throughput and fairness as a function of mobile user percentage.

and RateFair is caused by the "performance anomaly" of rate-based fairness criterion, where the sum-rate is limited by the user with the worst channel condition. Relaxed TimeFair gives the upper bound of the scheduling problem performance, which is the solution of the relaxed LP problem and is not achievable in practice due to the non-integer time slots it produces. However, the upper bound is well approximated by the TimeFair solution, indicating that our proposed algorithm achieves a near-optimal solution for the chosen communication sets. In absolute terms, the performance of TimeFair is quite good, being at least 2.5 times the performance of TDMA, which uses optimal SVD weights for each user but does not take advantage of parallelism among users.

A few other things are worth pointing out in Fig. 1. The first involves the performance-fairness tradeoff. Based on Fig. 1(a) and Fig. 1(b), this can be achieved by choosing the best operating point (ϵ) along the two curves. One use of this is to essentially solve the inverse optimization problem, namely to determine the best fairness possible for a given minimum performance threshold. Any other operating point in between the solutions to these two problems can also be determined from the plots. Finally, we note that GreedyMaxRate achieves very high sum-rate, as much as 4.7 times larger than TDMA and 67–100% higher than TimeFair. However, we also note that the minimal fairness constraint included in GreedyMaxRate means that it provides very low fairness, only around 0.3.

Fig. 2 shows the sum-rate and fairness index as a function of the mobile user percentage with $\epsilon = 0.1$. There are 30 users located randomly around the APs. The mobile user percentage varies from 0% to 25% and the scheduler is performed for all users. For mobile users with normal walking speed in indoor environment, the channel coherence time is set to 40 ms [12]. The expected value is calculated by assuming the channels of all users are static within the scheduling period, while the supported value is the achievable result taking into account the channel variation of mobile users. The achievable sum-rate and overall fairness degrade as more users become mobile, where the sum-rate loss is about 11% with 20% mobile users. Note that the change of mobile users' channels will not affect the static users' performance in the downlink case, since the precoders of mobile users are unchanged. Therefore, the sum-rate and fairness loss only come from the performance degradation of mobile users.

VI. Conclusion

In this paper, we presented a link scheduling algorithm for a cluster of multiple cooperative APs. The algorithm works in two phases: first, high-performing communication sets are generated via a weighted max sum-rate procedure, and then an integer programming problem is solved through relaxation and rounding to produce a schedule that provides near-optimal performance for the chosen communication sets and given fairness constraint. Although aggregate performance of our approach remains good with a minority of mobile users, fairness is reduced with the mobile users experiencing a performance loss. More work is necessary to investigate scheduling algorithms that can maintain good performance for mobile users and achieve strong fairness guarantees within the considered network scenarios.

References

[1] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, 2010.

[2] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance anomaly of 802.11b," in *Proc. IEEE INFOCOM 2003*, vol. 2, 2003, pp. 836–843.

[3] G. Tan and J. Guttag, "Time-based fairness improves performance in multi-rate WLANs," in *Proc. USENIX Conf.*, 2004, pp. 269–282.

[4] D. Blough, G. Resta, and P. Santi, "Interference-aware proportional fairness for multi-rate wireless networks," in *Proc. IEEE INFOCOM*, 2014, pp. 2733–2741.

[5] L. Cortes-Pena and D. Blough, "MIMO link scheduling for interference suppression in dense wireless networks," in *Proc. IEEE WCNC*, March 2015, pp. 1225–1230.

[6] F. Negro, S. Shenoy, I. Ghauri, and D. Slock, "On the MIMO interference channel," in *Proc. Inf. Theory Appl. Workshop*, 2010, pp. 1–9.

[7] M. Ge, J. Barry, and D. Blough, "Combined user selection and MIMO weight calculation for access point cooperation in dense wireless networks," *To appear in Proc. WCNC 2017.*

[8] S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, 2008.

[9] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted mmse approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, 2011.

[10] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, 2001.

[11] L. Cortes-Pena, J. Barry, and D. Blough, "Jointly optimizing stream allocation, beamforming and combining weights for the MIMO interference channel," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2245–2256, 2015.

[12] W. L. Shen, K. C. J. Lin, M. S. Chen, and K. Tan, "Sieve: Scalable user grouping for large MU-MIMO systems," in *2015 IEEE INFOCOM*, April 2015, pp. 1975–1983.