Opportunistic Fair Scheduling over Multiple Wireless Channels

Yonghe Liu and Edward Knightly
Department of Electrical and Computer Engineering
Rice University
Houston, TX 77005, USA
Email: {yhliu,knightly}@ece.rice.edu

***Abstract*— Emerging spread spectrum high-speed data networks utilize multiple channels via orthogonal codes or frequency-hopping patterns such that multiple users can transmit concurrently. In this paper, we develop a framework for opportunistic scheduling over multiple wireless channels. With a realistic channel model, any subset of users can be selected for data transmission at any time, albeit with different throughputs and system resource requirements. We first transform selection of the best users and rates from a complex general optimization problem into a decoupled and tractable formulation: a multi-user scheduling problem that maximizes total system throughput and a control-update problem that ensures long-term deterministic or probabilistic fairness constraints. We then design and evaluate practical schedulers that approximate these objectives.**

***Index Terms*— Weighted fair scheduling, multi-channel scheduling, probabilistic fairness guarantees, wireless networks**

I. INTRODUCTION

Achieving fair bandwidth allocation is an important goal for future wireless networks and has been a topic of intense recent research (see [1] for example and the references therein). In particular, in error-prone wireless links with a binary channel model (0 or 100% link error) it is impractical to guarantee identical *throughputs* to each user over short time scales; yet, over longer time scales, as channel conditions vary, lagging flows can "catch up" to re-normalize each flow's cumulative service (see [2] for example). Under a more realistic "continuous" channel model, any user can transmit at any time, yet users will attain different performance levels (e.g., throughput) and require different system resources depending on their current channel condition. Several scheduling algorithms have been designed for continuous channels that provide temporal or throughput fairness guarantees [3], [4], [5].

Regardless of the channel model employed, a common assumption of existing designs is that only a single user can access the channel at a given time, i.e., time division multiple access (TDMA). However, spread spectrum techniques are increasingly being deployed to allow multiple data users to transmit simultaneously on a relatively small number of separate high-rate channels. In particular, multiple logical channels can be created via different frequency hopping patterns or via orthogonal codes in Code Division Multiple Access (CDMA) systems. In this paper, we develop a novel framework for design and analysis of multi-channel wireless schedulers that opportunistically exploit variations in user channel conditions to select the best *set* of users and rates to schedule at each time instance subject to fairness and resource constraints. Our approach is as follows.

First, we develop a general methodology for design of opportunistic fair wireless schedulers using an adaptive control framework. The framework consists of a scheduling-optimization problem that guarantees a throughput-optimal selection of users and a control-parameter-update problem that ensures that the fairness constraints are satisfied. By decoupling the problem into two parts, the two sub-problems can be solved separately thereby significantly simplifying and standardizing the design procedure.

With this methodology, we next formulate and solve the multi-channel scheduling problem. Our key technique is to *jointly* exploit the temporal variations in the resource consumption of multiple users to opportunistically select users with greater throughput potential, while also ensuring that fairness constraints are satisfied. Towards this end, we develop MFS-D and MFS-P, a Multi-channel Fair Scheduler with Deterministic and Probabilistic fairness constraints respectively. For MFS-D the (long-term) expected throughput of different users are required to be equal, whereas for MFS-P they can differ by a fixed amount with a bounded probability. The two schedulers provide system operators with the flexibility to trade between stringency of the fairness guarantee and total system throughput. In both cases, we employ stochastic approximation based algorithms for updating the control parameter to ensure the respective fairness constraints are satisfied.

Finally, we perform an extensive set of simulations to evaluate the performance of MFS-D and MFS-P under channel models that incorporate mobility and fast fading. The simulations quantify the throughput gains of multi-channel scheduling under increasingly relaxed fairness constraints. Moreover, we study the impact of channel heterogeneity and system constraints such as limits on total power transmission.

The remainder of this paper is organized as follows. In Section II, we describe the framework for wireless scheduling. In Section III, we formulate the problem of scheduling multiple users concurrently on the wireless medium. In Section IV we develop MFS-D and in Section III MFS-P. Next, in Section VI we present simulation results. Finally, we review related

This research is supported by NSF Grants ANI-9979465 and ANI-0085842, a Texas Advanced Technology Program Grant, a Sloan Fellowship, and by gifts from Intel and Texas Instruments.

work in Section VII and conclude in Section VIII.

II. A FRAMEWORK FOR WIRELESS SCHEDULING

A. Preliminaries

We consider scheduling for a wireless system accessed by multiple users in which a centralized scheduler at the base station controls downlink scheduling, and uplink scheduling uses an additional mechanism such as polling to collect transmission requests from mobile nodes. We assume that downlink and uplink transmissions do not interfere with each other as in GPRS [6] for example.

Changing channel conditions are related to three basic phenomena: fast fading on the order of msec, shadow fading on the order of tens to hundreds of msec, and finally, long-time-scale variations due to user mobility. As our algorithm will exploit the users' channel conditions in making the scheduling decision, we consider systems with mechanisms to make predicted channel conditions available to the base station as is commonly the case with technologies such as HDR [7], UMTS-HS-DPA [8], (E)GPRS [6], etc.

To develop a scheduling algorithm that is applicable to a broad class of standards and systems, we abstract a user's channel condition into its "resource consumption", which reflects the *system* efficiency due to selecting that particular user for data transmission. For example, scheduling a user that currently has a poor quality channel would require consuming additional resources such as transmission power, stronger forward error protection, or longer transmission time due to lower data rates.

Due to inherent limits on the total system resources (e.g., power or time), high resource consumption by one user may prevent other users from being scheduled. In this way, resource consumption differs from "utility", as the former represents a cost to other users and the latter a gain to one particular user. We remark that resource consumption is a non-negative and non-increasing function of the channel quality indication (e.g., SNR).

B. Scheduler Design

Our objective in wireless scheduler design is to ensure fairness while simultaneously employing opportunistic scheduling strategies to increase the total system throughput by selecting users with high-quality channels when possible. To solve the problem, we observe that the two conflicting goals (throughput optimization and fairness guarantees) can be decoupled and solved as two separate entities as described in Figure 1: the control parameter updating block for fairness guarantees and the scheduling decision block targeting system throughput optimization.

A detailed description of the framework is as follows. The scheduling block makes scheduling decisions based on the channel condition $\vec{c}(k) = [c_1(k), \cdots, c_N(k)]$ and control parameters $\vec{w}(k) = [w_1(k), \cdots, w_N(k)]$. The output of the scheduling block is the scheduling decision $\vec{X}(k) = [X_1(k), \cdots, X_N(k)]$ for time slot k. In general, the scheduling decision represents a set of rates for simultaneous transmission of multiple users. That is, $X_i(k)$ denotes the scheduler's

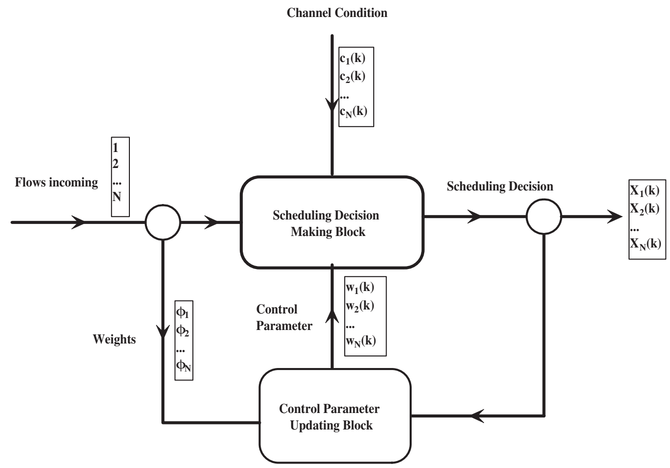


Fig. 1. A General Wireless Scheduling Formulation

selected transmission rate of user i in slot k. For the degenerate case of single-channel scheduling, the decision is simply the next flow to be serviced, e.g., $[0, \cdots, 1, 0]$. Selection of this decision vector is accomplished by solving an optimization problem, implicitly or explicitly, to maximize the system throughput or a function of it. Unlike the update block, the scheduling decision is temporally local in the sense that it only optimizes the objective at the current slot. This scheduling decision is output to the physical layer which transmits packets accordingly.

TABLE I
NOTATION SUMMARY

Term	Definition
$c_i(k)$	Channel condition of user i in time slot k
$w_i(k)$	Control parameter of user i in time slot k
$X_i(k)$	Scheduling decision for user i in time slot k
ϕ_i	Assigned weight of flow i

Because adapting to the channel condition will easily lead to short-term deviations from ideal fairness, memory of the decision history is required. Therefore, to enable service compensation at a later and more opportune time to under-serviced flows, the scheduling decision is fed back to the updating block. The functionality of the updating block is to update the control parameter $\vec{w}(k)$ to $\vec{w}(k+1)$ in such a manner that the output of the scheduling block will satisfy the fairness criteria on a larger time scale.

Observe that if a *fixed* optimal control parameter vector is given, the scheduling block will maximize the system throughput subject to the resource constraint. By monitoring the output of the scheduling block, the updating block ensures that the control parameter vector converges to the optimal value subject to the fairness constraint so that the overall system is optimized.

Thus, the philosophy of decoupling the system design into a scheduling block and control-update block is to simplify and structure the design process into two standard procedures, allowing different combinations of optimization objectives and fairness constraints. We next apply this design methodology to the multi-channel scheduling problem.

III. MULTI-CHANNEL PROBLEM FORMULATION

A. System Model

Using CDMA as an example, we formulate the fair scheduling problem over a multi-channel wireless system. In voice CDMA system, wireless users are separated via orthogonal codes with power control employed to maintain a particular Signal to Interference Noise Ratio (SINR) at the receivers [9]. In data CDMA systems, a number of higher-rate orthogonal channels are available for data transmission (typically fewer than the number of users). The multi-channel scheduling problem is to select the times, channels, and rates for transmission of queued packets (note that users can receive on any pre-assigned code, albeit with different resource consumption). In this paper, total transmission power is considered to be the system resource constraint while the power requirement per bit is used as an indication of a user's channel condition. Following the system model for scheduling described in Section II, we present the following multi-channel formulation.

Consider N users accessing the system such that user i has a set of possible transmitting rates in slot k given by $R_i(k) \in \{0, R_i^1, ..., R_i^{M_i}\}$, where $(M_i + 1)$ denotes the number of the possible rates for user i, and rate 0 indicates that the user is not scheduled at that time. Such a rate set is enabled by, for example, using different spreading codes and/or modulation schemes on the different wireless channels. In time slot k, user i experiences a certain wireless channel condition $c_i(k)$ abstracted as a per bit power consumption in order to guarantee a certain SINR. In other words, the transmitting user i in slot k using rate $R_i(k)$ requires power $C_i(k) = c_i(k)R_i(k)$.[1] Notice that $c_i(k)$ is a random process reflecting the user's channel condition as driven by user mobility and channel shadowing.

B. Objective and Resource Constraint

The objective of the scheduler is to maximize the system throughput subject to the fairness and resource constraints. Mathematically, let $X_i(k)$ denote the transmission rate of user i in time slot k and let $Y(k) = \sum_{i=1}^{N} X_i(k)$ denote the total throughput in slot k. The objective of the scheduler is then to maximize the expectation of $Y(k)$, i.e.,

$$\max \qquad Y = E(\sum_{i=1}^{N} X_i(k)) \tag{1}$$

One constraint is the total system resource limitation which in a CDMA system is the maximum transmission power limit in each time slot given by

$$\sum_{i=1}^{N} c_i(k)X_i(k) \le P, \tag{2}$$

where P denotes the maximum total power transmission regulated.

[1]We assume a linear relationship of the scheduled rate and power consumption as applicable to (for example) transport format selection in a UMTS system [8]. A non-linear relationship is a trivial but lengthy extension.

The second constraint is fairness, and we consider two fairness objectives: deterministic and probabilistic. The two fairness criteria lead to two scheduler designs described in Section IV and V and provide techniques for network operators to trade stricter fairness for higher throughput while in both cases maintaining quantifiable fairness characteristics at longer time scales.

IV. SCHEDULER DESIGN FOR DETERMINISTIC FAIRNESS

In this section, we devise MFS-D, a multi-channel scheduler for deterministic fairness. Let $Y_i = E(X_i(k))$ denote the expected throughput for user i. Given user i's target weight ϕ_i in the system, deterministic fairness requires

$$\frac{Y_i}{\phi_i} = \frac{Y_j}{\phi_j}. \tag{3}$$

That is, for any two flows i and j, their expected throughput should be exactly proportional to their assigned weights. Therefore the scheduling decision can be formulated as the following optimization problem.

$$\max \qquad Y = \sum_{i=1}^{N} Y_i \tag{4}$$

$$s.t.$$

$$\frac{Y_i}{\phi_i} = \frac{Y_j}{\phi_j} \tag{5}$$

$$\sum_{i=1}^{N} c_i(k)X_i(k) \le P \tag{6}$$

$$X_i(k) \prod_{l=1}^{M_i} (X_i(k) - R_i^l) = 0, \; 1 \le i \le N \tag{7}$$

To solve this problem, we observe that the above is equivalent to the following problem.[2]

$$\max \qquad Z \tag{8}$$

$$s.t.$$

$$Z \le \frac{Y_i}{\phi_i} \;, \; 1 \le i \le N \tag{9}$$

$$\sum_{i=1}^{N} c_i(k)X_i(k) \le P \tag{10}$$

$$X_i(k) \prod_{l=1}^{M_i} (X_i(k) - R_i^l) = 0 \;, \; 1 \le i \le N \tag{11}$$

Intuitively, if we maximize the minimum $\frac{Y_i}{\phi_i}$, the system throughput is maximized subject to the fairness constraint. Consequently, the first constraint leads to the requirement that

$$Z = \frac{Y_i}{\phi_i} \quad ,1 \le i \le N, \tag{12}$$

since one can easily reduce the throughput for users with surplus. Moreover, since each $\frac{Y_i}{\phi_i}$ is identical, the objective function is equivalent to maximizing

$$Z = \sum_{i=1}^{N} w_i Y_i, \quad w_i \ge 0 \tag{13}$$

[2]See [3], [10] for other applications of such a transformation technique.

where w_i is a non-negative constant. Thus, if (1) we can maximize Z defined in (13) for a fixed $\vec{w}^*$, and (2) we can also satisfy the fairness constraint, the resulting scheduler will be the optimal solution to the original problem.

We remark that a fixed control vector leading to the optimal solution is not available, as the channel information can not be obtained in advance, and the channel conditions may change over time. Thus, as described in Section II, we dynamically adjust the control parameters online such that it converges to the optimal solution. Next, based on the above discussion, we introduce a control parameter $w_i(k)$ for each user i in time slot k and design the scheduling block and updating block respectively.

A. Design of the Scheduling Block

According to the framework, the functionality of the scheduling block is to ensure a throughput-optimal selection of users and rates. According to (13), the objective of the scheduling block is set to

$$\max \quad \sum_{i=1}^{N} w_i(k)X_i(k), \tag{14}$$

such that in each time slot the weighted system throughput is maximized. The constraints of this optimization are given by the following.

$$\sum_{i=1}^{N} c_i(k)X_i(k) \leq P \tag{15}$$

$$X_i(k)\prod_{l=1}^{M_i}(X_i(k) - R_i^l) = 0 \quad ,1 \leq i \leq N \tag{16}$$

Notice that there is no fairness constraint as (from Section II) fairness is treated in the updating block.

We observe that this optimal scheduling block formulation is an NP hard Knapsack problem [11]. Consequently, since a complete search of the solution space is infeasible in practice, we develop an approximation to the optimal solution as follows. Observe that the larger the ratio $\frac{w_i(k)}{c_i(k)}$, the more likely the user should be selected, as it adds increasingly to the objective function. Hence a greedy algorithm to solve this problem can be formulated by first generating the following sorted list:

$$\frac{c_1(k)}{w_1(k)} \leq \frac{c_2(k)}{w_2(k)} \leq ... \leq \frac{c_N(k)}{w_N(k)}. \tag{17}$$

Then select $X_i = \max(0, R_i^1, ..., R_i^{M_i})$ beginning with ordered-flow 1 and proceeding sequentially until flow j such that the maximum power limit is reached. (Notice that flow j may have to select a rate smaller than its maximum possible.) For all $i > j$, the transmission rate in time slot k will be set to zero. This approximation to the optimal solution has computational complexity due to the sort of $O(Nlog(N))$.

However, the deviation of the greedy algorithm from the optimal solution is unbounded. For example, assume 2 users both with rate set $\{0, 1\}$ are in the system. Let $w_1(k) = 10$, $w_2(k) = 10P - 1$, $c_1(k) = 1$, and $c_2(k) = P$, where P is the total transmission power. According to the greedy algorithm, only user 1 will be selected as long as P is greater than 2. In contrast, the optimal solution is to select user 2 so that the deviation between the two solutions increases with P.

A simple enhancement to limit the deviation from the optimal solution is as follows. First, still employing the greedy algorithm, we generate user sets $T = \{1, 2, ..., j\}$ and $T' = \{(j+1)\}$ where T is the solution given by the greedy algorithm and $j, (j + 1)$ are the index of the sorted list described as above. If set T generates a better result than T', T is selected; otherwise, T' is selected. This algorithm has been shown to achieve an objective within a factor of 2 to the optimal value. Readers can find other enhanced approximation algorithms in [11] at the cost of increased computational complexity.

We make two observations about the scheduling block. First, note that we consider the resource limits and rate set as constraints. If the system has other constraints, they can be added into the scheduling block optimization problem. For example, if user i and user j can not be scheduled simultaneously in one time slot, we can set another constraint $X_i(k)X_j(k) = 0$. This may be important for particular underlying physical layer implementation. For example, in a CDMA network, two users may not be assigned in the same time slot in order to maintain code orthogonality.

Second, observe that although a rate set is assigned to each user, the maximum rate is most often used in the approximation to the Knapsack problem. With a linear relationship between $c_i(k)$ and $X_i(k)$, other rates are only used for the last selected user to utilize the remaining total transmission power. However, we remark that for a non-linear relationship, the rate set will be of increased importance in selecting users.

B. Design of the Updating Block

The function of the updating block is to update the control parameters in order for the output of the scheduling block to satisfy the fairness requirement. Recall that the requirement of the updating block is to guarantee that the control vector converges to the optimal value while satisfying the fairness constraint. To ensure this, we employ the stochastic approximation algorithm in the updating block. We note that such a technique is first employed in [4] in the context of single channel scheduling.

Define a vector function $f(\vec{w}) = [f_1(\vec{w}), ..., f_N(\vec{w})]$, where

$$f_i(\vec{w}) = \frac{\phi_i}{\sum_{j=1}^{N}\phi_j} - E(\frac{X_i(k)}{\sum_{j=1}^{N}X_j(k)}), \tag{18}$$

and $\vec{w} = [w_i, ..., w_N]$ denotes the adaptive control vector. The deterministic fairness constraint of Equation (3) is equivalent to the requirement $f(\vec{w}) = 0$. In other words, the control parameter updating algorithm has to find the solution $\vec{w}^*$ for this function. Notice that the value of $f_i(\vec{w})$ depends on the output of the scheduler decision so that the updating algorithm should be dynamically adjusted according to scheduling decisions.

Stochastic approximation is an effective technique for finding zeros of a function $f(\cdot)$ which cannot be explicitly known [12]. If a noisy measurement is available, i.e., $y^k = f(x^k)+e^k$, where e^k denotes the observed noise, stochastic approximation recursively estimates the root for $f(\cdot)$ by

$$x^{k+1} = x^k - a^k y^k, \tag{19}$$

where a^k is the step size. If e^k is white noise and a^k converges to zero, x^k will converge with probability 1 to the root of $f(\cdot)$ if certain conditions on $f(\cdot)$ are satisfied.[3]

In our case, since $f(\vec{w})$ is an expected value, we can not directly observe it. However, we do have a noisy observation $y^k = [y_1^k, ..., y_N^k]$, where

$$y_i^k = \frac{\phi_i}{\sum_{j=1}^N \phi_j} - \frac{X_i(k)}{\sum_{j=1}^N X_j(k)} \tag{20}$$

The expected value of the observation error in this case is

$$E(e^k) = E(y^k - f(w^k)) = 0. \tag{21}$$

Therefore, we can use the stochastic approximation algorithm to adaptively find $\vec{w}^*$ as

$$w_i(k+1) = w_i(k) - a^k y_i^k, \tag{22}$$

where a^k is chosen to converge to zero, e.g., $a^k = 1/k$.

Thus, the resulting scheduler MFS-D contains the above scheduling and update blocks. Note that this solution maximizes the total normalized throughput in each time slot, ensures that $\vec{w}$ converges to a fixed $\vec{w}^*$, and ensures that the deterministic fairness constraint is satisfied. Thus, MFS-D provides an approximate solution to the optimization problem of Equations (4) to (7).

V. SCHEDULER DESIGN FOR PROBABILISTIC FAIRNESS

In this section, we use the methodology of Section II to develop MFS-P, a multi-channel scheduler for probabilistic fairness.

In [5] a wireless scheduler is designed under a probabilistic fairness index defined by

$$Pr(\left|\frac{Y_i}{\phi_i} - \frac{Y_j}{\phi_j}\right| > x) \le f(i,j,x) \tag{23}$$

Notice that Equation (23) defines probabilistic fairness on the entire distribution of the service difference between user i and user j. To simplify the scheduler design, we relax the probability constraint to a fixed point. In other words, for the multiple channel scheduling problem, we use the following fairness constraint to replace (23)

$$Pr(\left|\frac{Y_i}{\phi_i} - \frac{Y_j}{\phi_j}\right| > \delta) \le P_\delta, \tag{24}$$

where P_δ the targeted probability and δ is an operator-specified constant denoting the service discrepancy where probabilistic fairness is to be ensured. For a fixed P_δ, a lower value of δ provides for more stringent fairness such that $\delta = 0$ provides that two users' throughputs have 0 difference with probability of no more than P_δ. With a scale factor of $\sum_{i=1}^N Y_i$ on δ, this equation can be further transformed into

$$Pr(\left|\frac{Y_i}{\sum_{j=1}^N Y_j} - \phi_i\right| > 0.5\delta\phi_i) \le P_\delta \tag{25}$$

[3]Different stochastic algorithms require different conditions. General requirements include stationarity and a certain order differential. Readers are referred to [12] for a detailed discussion.

A. Design of the Scheduling Block

We employ the same control parameter $w_i(k)$ as in MFS-D and formulate the objective function of the scheduling block as

$$\max \quad \sum_{i=1}^N w_i(k) X_i(k). \tag{26}$$

Together with the resource and rate constraints, this again is a Knapsack problem such that MFS-P's scheduling block employs the same algorithmic solution as in Section IV.

A general proof of optimality of the MFS-P scheduling block is difficult to obtain due to the probabilistic fairness constraint. However, for the special case of $\delta = P_\delta = 0$, the proof for MFS-P is identical to that of MFS-D. However, in general as well as in this special case, the actual scheduling decisions and control updates for MFS-P and MFS-D are quite different, as quantified in Section VI.

B. Design of the Updating Block

We employ stochastic approximation to ensure the probabilistic fairness constraint via convergence of control parameter updating as follows. Define a vector function of the control parameter $(\vec{w})$ as

$$f(\vec{w}) = [f(w_1), ..., f(w_i), ..., f(w_n)] \tag{27}$$

where

$$\begin{aligned} f(w_i) &= Pr(\left|\frac{Y_i}{\sum_{j=1}^n Y_j} - \phi_i\right| > 0.5\delta\phi_i) - P_\delta \\ &= E(I(\left|\frac{Y_i}{\sum_{j=1}^n Y_j} - \phi_i\right| > 0.5\delta\phi_i)) - P_\delta. \end{aligned}$$

The target is to have $f(\vec{w}) = 0$. Notice that $f(\vec{w})$ is an expected value and is not directly observable. However, in each time slot k, we have a noisy observation

$$I(\left|\frac{Y_i(k)}{\sum_{j=1}^n Y_j(k)} - \phi_i\right| > 0.5\delta\phi_i) - P_\delta, \tag{28}$$

where $Y_i(k)$ is the user i throughput at time slot k. A measured and smoothed version of $Y_i(k)$ can be obtained as

$$Y_i(k) = (1-\beta)Y_i(k-1) + \beta X_i(k), \tag{29}$$

where β is the filter parameter.

Therefore, we can use the stochastic approximation algorithm to update the control parameters again as follows.

$$w_i(k+1) = \tag{30}$$
$$\begin{cases} w_i(k) + \alpha(1-P_\delta) & (\frac{Y_i}{\sum_{j=1}^N Y_j} - \phi_i < -0.5\delta\phi_i) \\ w_i(k) & (\left|\frac{Y_i}{\sum_{j=1}^N Y_j} - \phi_i\right| \le 0.5\delta\phi_i) \\ w_i(k) - \alpha(1-P_\delta) & (\frac{Y_i}{\sum_{j=1}^N Y_j} - \phi_i > 0.5\delta\phi_i) \end{cases}$$

We remark that in (28), the relationship between $(\frac{Y_i(k)}{\sum_{j=1}^n Y_j(k)} - \phi_i)$ and $0.5\delta\phi_i$ is only captured as an indicator function which takes on values 0 and 1. This is rooted from the fairness requirement measured at only one service discrepancy, δ. While a simple updating scheme, this approach may have larger unfairness for deviations larger than δ. To limit such unfairness, the updating algorithm can be refined as follows.

$$w_i(k+1) = \qquad (31)$$
$$\begin{cases} w_i(k) - \alpha(1-P_\delta)(\frac{Y_i(k)}{\sum_{j=1}^{n} Y_j(k)} - \phi_i + 0.5\delta\phi_i) \\ \qquad (\frac{Y_i}{\sum_{j=1}^{N} Y_j} - \phi_i < -0.5\delta\phi_i) \\ w_i(k) \\ \qquad (|\frac{Y_i}{\sum_{j=1}^{N} Y_j} - \phi_i| \le 0.5\delta\phi_i) \\ w_i(k) - \alpha(1-P_\delta)(\frac{Y_i(k)}{\sum_{j=1}^{n} Y_j(k)} - \phi_i - 0.5\delta\phi_i) \\ \qquad (\frac{Y_i}{\sum_{j=1}^{N} Y_j} - \phi_i > 0.5\delta\phi_i) \end{cases}$$

The added terms measure the deviation of the current fairness from the requirements and scale the adjustment speed on the control parameters. This leads to faster convergence and less fluctuation.

VI. Simulation Experiments

In this section, we present an extensive set of simulation experiments to evaluate the performance of the deterministic and probabilistic multichannel fair schedulers (MFS-D and MFS-P) designed in Sections IV and V.

Our experimental design considers throughput and fairness as the primary performance measures and we consider a number of factors of scheduler performance. We first compare the dynamic behavior of MFS-P and MFS-D in terms of both control parameter update and short-term fairness behavior. We next study the fundamental throughput-fairness tradeoff by exploring the impact of the operator designated fairness discrepancy parameter δ on fairness and total system throughput. Third, we study the effects of scaling the number of users in the system in multi-channel vs. single channel schedulers. Next, we consider the effect on fairness of one user having a perpetually good channel. Finally, we explore the impact of total system power on throughput and fairness.

A. Channel Model

To explore the role of the channel conditions on system throughput and per-user fairness, we consider a simple model to capture the effects of mobility and fading. In the model, power consumption per bit ranges from 0 (best) to 1 (worst). The channel condition is represented by a random process consisting of a sinusoid with random phase plus additive noise. That is, the channel condition for user i at time t is given by

$$c_i(t) = 0.5 + d\cos(2\pi f_i t + \theta_i) + X_{\sigma_i}(t) \qquad (32)$$

where $\theta_1, \theta_2, \cdots$ are independent and uniformly distributed in $[0, 2\pi)$ giving the channel conditions statistically independent phases.

The sinusoidal term represents the long time scale effects of mobility for different mobility speeds and channel time scales $1/f_i$ and d represents the range of the channel effects due to mobility. The additive noise $X_{\sigma_i}(t)$ represents a model of the effects of Rayleigh and shadow fading via the conservative assumption of additive white Gaussian noise with variance σ_i^2. This model allows us to study the influence of the experienced channel on both system throughput and fairness.

We assume that power control is perfect and the transmission rate equals the throughput at each time slot. The power consumption for user i transmitting at rate $R_i(k)$ in slot k is simply $c_i(k)R_i(k)$. All users in the system are allocated with the rate set $\{0, 1\}$. We consider traffic in which all flows are continuously backlogged such that the achieved fairness and throughput is entirely related to the scheduling process and channel conditions without any variation due to traffic fluctuations. Moreover, we assume that data can be dynamically fragmented to fit into one time slot at the specified transmission rate. Unless otherwise specified, $f_i = 0.005$, $d = 0.3$, and $\sigma_i = 0.2$.

B. Scheduler Dynamics

A key distinction between MFS-D and MFS-P is their control parameter updating blocks that target different fairness guarantees. Intuitively, the updating block of MFS-P is less reactive as the control parameter is updated only when the threshold of discrepancy is triggered. In contrast, MFS-D updates the control parameter at every time slot to satisfy the stricter fairness constraint.

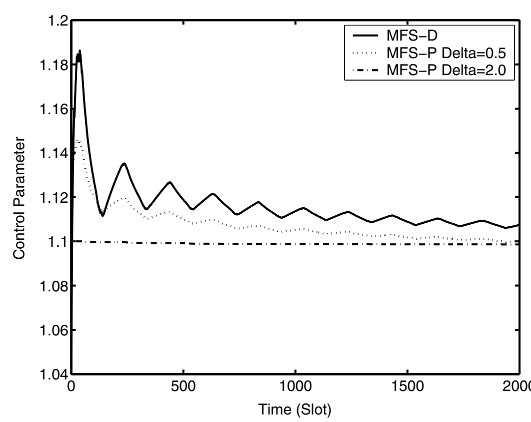


Fig. 2. MFS-D and MFS-P Control Parameter Update

This phenomena is illustrated in Figure 2 which depicts a temporal snapshot of the control parameter updating procedures for MFS-D and MFS-P. Notice that in MFS-D, the user's control parameter is continuously being updated due to its stricter fairness requirement. The oscillating pattern of the updating curve represents the scheduler's effort to track and converge to the underlying channel condition. In order to maintain identical normalized throughput in the long term, MFS-D has to accumulate "credits" via control parameter updating for bad channel users in order for it to acquire a transmission opportunity in the future.

In contrast, control parameter updates for MFS-P occur more smoothly and the control parameter stays constant for a certain time interval, especially for a larger δ. The means that during these slots, the condition $(|\frac{Y_i}{\sum_{j=1}^{N} Y_j} - \phi_i| \le 0.5c\phi_i)$ is satisfied and the updating procedure is not triggered. For the special case of $\delta = 2$ and above, a user's control parameter is not increased even if its throughput is zero, and a continuously bad channel user will not accumulate credit for catching up at a later time. Thus, as further elaborated below, $\delta < 2$ is required to ensure fairness.

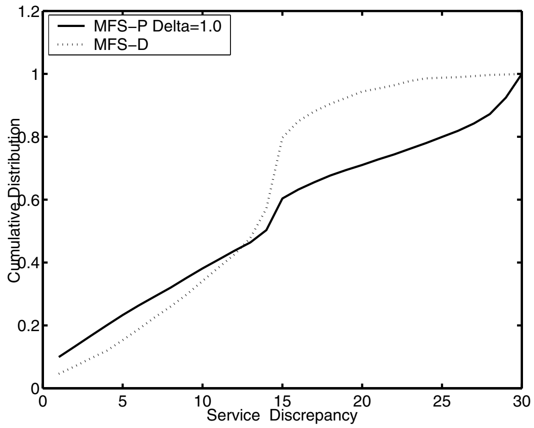


Fig. 3. Distribution of Short Term Unfairness

The less stringent update policy of MFS-P also has an impact on short-term fairness. In particular, with MFS-P a bad channel user can temporarily lose its chance for transmission until the control parameter threshold is met. Figure 3 depicts a histogram of the normalized service discrepancy between two users with a 30 time slot measurement interval. Observe that MFS-P has more density at high discrepancies indicating that MFS-P allows the scheduler to deviate farther from perfect fairness than does MFS-D.

C. Throughput and Fairness

The goal of opportunistic scheduling is throughput maximization subject to the fairness guarantee and resource constraints. Here, we study the throughput gain of MFS-P as a function of the fairness threshold parameter δ. From Equation (25), the parameter δ represents the normalized service discrepancy that is allowed with probability P_δ. For example, with $\delta = 1$ and identical weights, two flow's throughputs can differ by up to a factor of 1 with probability P_δ. Similarly, with $\delta = 0$, any two flows' service is maintained to be identical with probability P_δ.

In this set of simulations, 16 users are active with 12 users having weight 1 and 4 users having weight 2.[4] The maximum transmission power is set to be 2 and $P_\delta = 0.2$. The total simulation time is 2000 slots.

Figure 4 depicts the effect of the parameter δ on throughput. The x-axis depicts δ ranging from 0 to 3 with step length 0.5. Each bar in the graph contains three components: the lower bar represents the average throughput per time slot normalized to the weight unit, the middle and upper bars respectively depict the standard deviation and the maximum difference between any two users' normalized throughput. Essentially, the lower bar represents the efficiency of the scheduler in increasing system throughput, the middle bar denotes the average system unfairness, and the upper bar represents the maximum system unfairness.

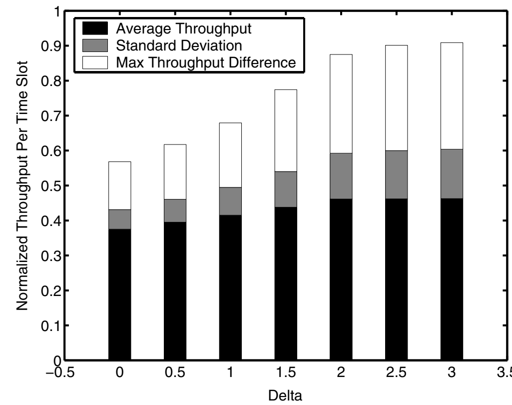


Fig. 4. Throughput vs. Fairness Parameter δ

Observe that system throughput increases roughly 25% with increasing δ, as larger values of δ represent increasingly relaxed fairness constraints. Moreover, notice that the throughput cannot increase further when δ is greater than 2: when $\delta \geq 2$, according to (31), a user's parameter will not be updated even if its throughput is zero. Between these extremes, a wide range of δ's yield an effective tradeoff between throughput gains and fairness.

D. Number of Users

A larger number of users provides an increased degree of freedom for the scheduler to exploit varying channel conditions and select the best subset of users for transmission. In this set of simulations, we investigate the relationship between the total system throughput and number of users for multi- and single-channel schedulers. For MFS-P, we consider δ given by 0.5, P_δ given by 0.2, and all users having weight 1.

Figure 5 shows the total system throughput per time slot versus the number of users in the system. The figure indicates that the total system throughput increases almost almost linearly with the number of users until saturation due to the system's total power constraint. This implies that with joint optimization of multiple channel scheduling over multiple users, the average throughput per user remains roughly unchanged with an increased number of users in the system, up to the fundamental throughput limits of the system due to physical constraints.

To explore this result more deeply and compare multi-channel to single-channel scheduling, we perform a set of simulations in which each user's channel conditions is uniformly distributed in [0,1] and the total transmission power is still 2. Moreover, to isolate the effects of the channel and user and rate selection, we fix the control parameter to always be 1 such that the best channel(s) are always selected without regard to fairness.

Figure 6(a) illustrates that this linear relationship holds for a wide range of maximum per-user rates from 1 to 11. Figure 6(b) depicts the result for *single-channel* scheduling under the same simulation setup and shows that the total system

[4] Use of different weights in this scenario gives more flexibility to the scheduler.

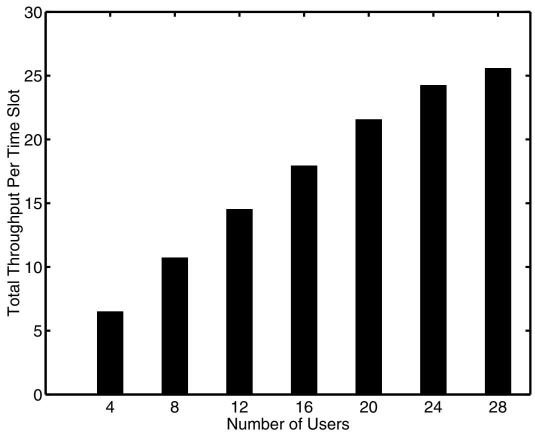
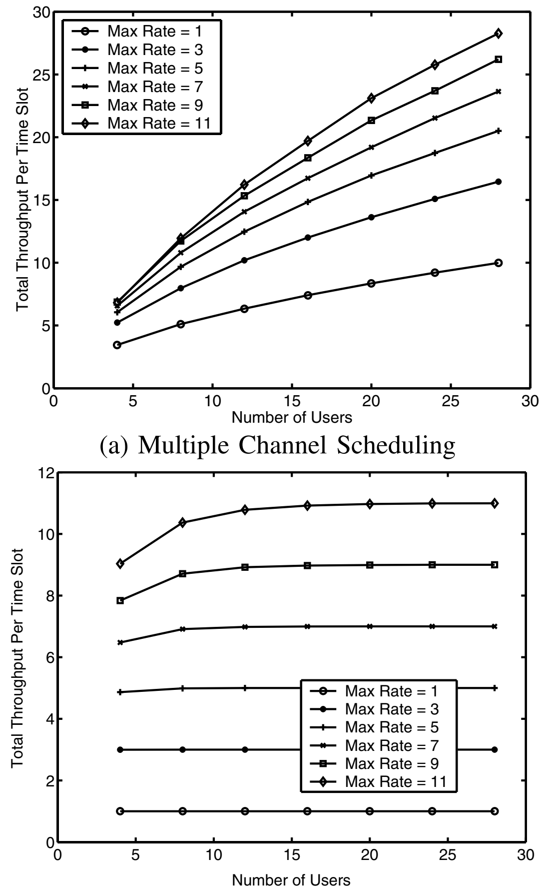


Fig. 5. Throughput vs. Number of Users



(a) Multiple Channel Scheduling

(b) Single Channel Scheduling

Fig. 6. Throughput vs. User Number

throughput is strongly dependent on the maximum user rate available in the system. When the maximum rate is low, single-channel scheduling is limited by the one-user-at-a-time rule and thus cannot claim unused power in the system. As the maximum rate increases, the total system throughput begins to increase with the number of users as well, although at a significantly lesser rate. Observe that equal throughput is achieved in both systems only in the extreme case in which the maximum user rate in the single-channel system is limited only by the total system power, an unrealistic scenario in cellular systems.

E. Heterogeneous Channel Conditions

Here, we consider the impact of users having perpetually different channel conditions. For MFS-D and MFS-P, we use an approximate solution to the Knapsack problem in the scheduling block. After scheduling users according to the algorithm described in Section IV-A, the scheduler checks the remaining users to determine if one can be fit in to to claim the excess power. If so, the user will be scheduled regardless of its control parameter or position in the sorted list. Observe that when a user has an extremely good channel and hence little resource consumption, the scheduler is likely to select this user using the remaining power after scheduling other users.

In these experiments, we modify the above simulation scenario such that one “best-channel” user has a *constant* high quality channel condition. In different simulations, this user’s power per-bit consumption varies from 0.01 to 0.3.

Figure 7 shows the results and depicts two bars for each channel condition. The right bar denotes the throughput for the user with the perpetually high-quality channel and the left bar denotes the average throughput of all other users. Observe that the best-user’s throughput decreases as its channel condition gets worse. However, when its channel condition is extremely good (the left-most bars), this user is scheduled in almost every time slot. The reason is that although this user continuously incurs a decreasing fairness control update parameter due to receiving high service rates, it can nearly always be fit into the remaining power budget which cannot be claimed by the other users with lower quality channels. Thus, while the attained throughputs are not strictly “fair”, it is indeed a desirable outcome as this throughput obtained by the best-channel user would otherwise be wasted. In other words, unique to multiple channel scheduling, unfairness may become not only unpreventable, but also desirable.

Similarly, Figure 8 illustrates the effects of having a single user with a perpetually poor channel condition. While the degradation in that user’s throughput is expected with increasingly poor channel conditions, the key observation is that the average system throughput is nearly unchanged. Thus, in the multi-channel system, other users are not forced to wait while the lagging bad-channel flow catches up, as such a user occupies at most one channel.

F. Maximum Transmission Power

The above issue of fitting in an “extra” best-channel user becomes more pronounced when the total resource is relatively large, as a better channel user can be more easily accommodated in the remaining power budget. This issue is also affected by the transmission rate set: in the limit, if a single user can utilize all available resources and transmit at a very high rate, no remaining power budget would be available to scheduler other users.

Figure 9 depicts the results of experiments that consider the effects of varying the total transmission power from 2 to 8. The

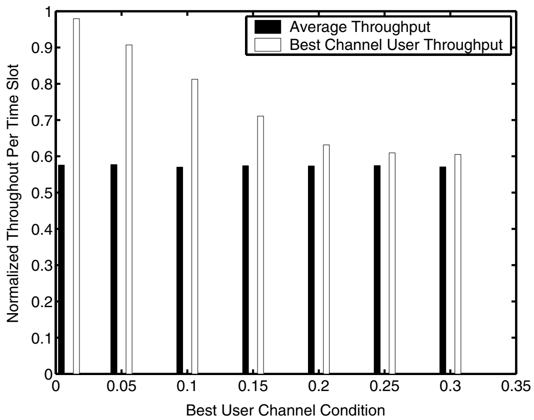


Fig. 7. Throughput vs. Best User's Channel Condition

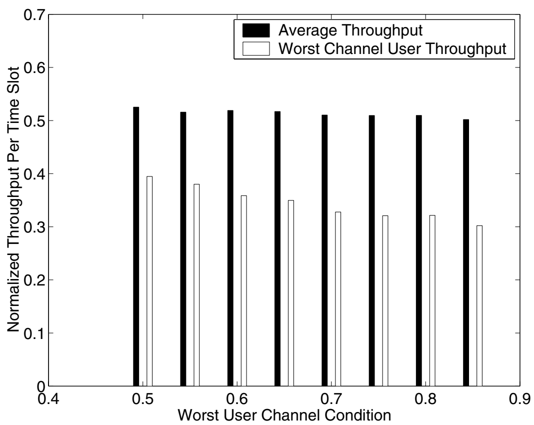


Fig. 8. Throughput vs. Worst User's Channel Condition

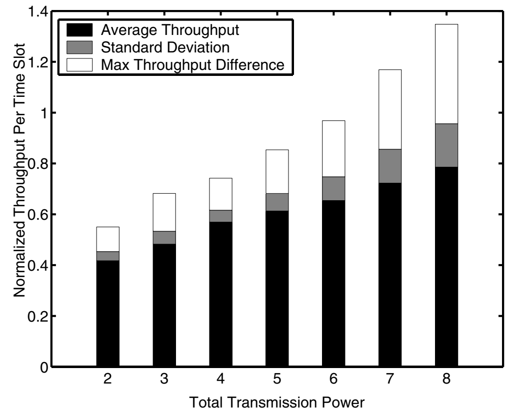


Fig. 9. Throughput vs. Maximum Total Transmission Power

figure illustrates that for the reasons described above combined with the flexibility of multi-channel scheduling, a larger total transmission power results in increases in the normalized average throughput, as well as the standard deviation and maximum difference in throughput.

VII. Related Work

While many fair wireless schedulers were designed under the binary channel model e.g., [1], a number of schedulers have been designed under a more realistic multi-rate channel model [3], [4], [5], [13], [14], [15], [16]. Such schemes have the advantage of being able to opportunistically exploit channel variations to select good-channel users while also satisfying fairness constraints. In contrast, MFS considers a more general multi-channel formulation applicable in spread spectrum networks in which multiple users can be scheduled simultaneously, again at multiple rates. As illustrated in Section VI, multi- and single-channel systems behave quite differently in terms of their achievable throughput and fairness characteristics.

At the physical layer, multiple-channel power control in CDMA networks is an area of intense study, e.g., [17], [18], [19], [20]. The objective of such research is to select the set of user transmission powers to optimally allocate the total power budget subject to per-user SINR requirements or per-user throughput requirements. However, such power allocation problems are quite different from our formulated multi-channel scheduling problem as the latter determines which users to schedule when and at which rates as given by the formulated optimization problem. In other words, in contrast to a purely physical layer optimization, MFS-D and MFS-P provide an optimal MAC-layer scheduling algorithm that exploits physical-layer information on channel conditions to dictate the rates, times, and powers of each user's transmission.

Finally, the generic problem of scheduling a set of jobs over multiple resources (machines) has been studied in [21], [22], [23] for example. Such papers focus on minimizing the average job delay and developing on-line algorithms that approximate the optimal solution with bounded error. However, such solutions are not applicable here as the multi-channel wireless scheduling problem has a unique formulation such as constraints of fairness among users, high variability in the resource availability and cost (i.e., a complex channel model), etc.

VIII. Conclusions

This paper formulates the problem of opportunistically scheduling multiple users concurrently in wireless networks. We introduced and analyzed Multi-channel Fair Scheduler (MFS), the first wireless scheduling algorithm that provide long term deterministic (MFS-D) and probabilistic (MFS-P) fairness guarantees respectively over multiple wireless channels. By considering resource consumption over different channels, the algorithms allow system operators to *jointly optimize* the transmission over multiple channels for total throughput maximization while maintaining flexible fairness constraints.

References

[1] V. Bharghavan, S. Lu, and T. Nandagopal, "Fair scheduling in wireless networks: Issues and approaches," *IEEE Personal Communications*, pp. 44–53, Feb. 1999.

[2] T. Ng, I. Stoica, and H. Zhang, "Packet fair queueing algorithms for wireless networks with location dependent errors," in *Proceedings of IEEE INFOCOM '98*, San Francisco, CA, Mar. 1998.

[3] S. Borst and P. Whiting, “Dynamic rate control algorithms for HDR throughput optimization,” in *Proceedings of IEEE INFOCOM ’01*, Anchorage, Alaska, Apr. 2001.
[4] X. Liu, E. Chong, and N. Shroff, “Transmission scheduling for efficient wireless network utilization,” in *Proceedings of IEEE INFOCOM ’01*, Anchorage, Alaska, Apr. 2001.
[5] Y. Liu, S. Gruhl, and E. Knightly, “WCFQ: an opportunistic scheudler with statistical fairness guarantee,” to appear in *IEEE Transactions on Wireless Communications*.
[6] Digital cellular telecommunications system (Phase 2+) European Telecommunications Standards Institute, “Overall description of the GPRS radio interface; stage 2 (GSM 03.64 version 8.5.0 release 1999,” http://www.etsi.org, Aug. 2000.
[7] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, “CDMA/HDR: a bandwidth efficient high speed wireless data service for nomadic users,” *IEEE Communications Magazine*, vol. 38, no. 7, pp. 70–77, July 2000.
[8] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, “Physical layer aspects of utra high speed downlink packet access (release 4) tr25.848 v4.0.0,” http://www.3gpp.org, Mar. 2001.
[9] T. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall, 1996.
[10] S. Kandukuri and S. Boyd, “Optimal power control in interference-limited fading wireless channels with outage-probability specifications,” *IEEE Transactions on Wireless Communications*, vol. 1, no. 1, pp. 46–55, Jan. 2002.
[11] S. Martello and P. Toth, *Knapsack Problems: Algorithms and computer Implementations*, John Wily and Sons, Ltd., New York, 1990.
[12] H. Kushner and D. Clark, *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Springer-Verlag, 1978.
[13] M. Andrews, S. Borst, F. Dominique, P. Jelenkovic, K. Kumaranand, K. Ramakrishnan, and P. Whiting, “Dynamic bandwidth allocation algorithms for high-speed data wireless networks,” *Bell Labs Technical Journal*, Jul.-Sep. 1998.
[14] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, “Providing quality of service over a shared wireless link,” *IEEE Communications Magazine*, pp. 150–154, Feb. 2001.
[15] G. Holland, N. Vaidya, and P. Bahl, “A rate-adaptive MAC protocol for multi-hop wireless networks,” in *Proceedings of ACM MOBICOM ’01*, Aug. 2001.
[16] B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly, “Opportunistic media access for multirate ad hoc networks,” in *Proceedings of ACM MOBICOM ’02*, Sept. 2002.
[17] K. Gilhousen, I. Jacobs, R. Padovani, A. Viterbi, L. Weaver, and C. Wheatley III, “On the capacity of a cellular CDMA system,” *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 303–312, Feb. 1991.
[18] J. Lee, R. Mazumdar, and N. Shroff, “Downlink power allocation for multi-class CDMA wireless networks,” in *Proceedings of IEEE INFOCOM ’02*, New York, June 2002.
[19] P. Marbach and R. Berry, “Downlink resource allocation and pricing for wireless networks,” in *Proceedings of IEEE INFOCOM ’02*, New York, June 2002.
[20] R. Yates, “A framework for uplink power control in cellular radio systems,” *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 1341–1347, Feb. 1995.
[21] J. Chen and A. Miranda, “A polynomial time approximation scheme for general multiprocessor job scheduling,” *SIAM Journal on Computing*, vol. 31, no. 1, pp. 1–17, May 2001.
[22] R. Li and L. Shi, “An on-line algorithm for some uniform processor scheduling,” *SIAM Journal on Computing*, vol. 27, no. 2, pp. 414–422, Apr. 1998.
[23] W. Schwiegelshohn, W. Ludwig, J. Wolf, J. Turek, and P. Yu, “Smart smart bounds for weighted response time scheduling,” *SIAM Journal on Computing*, vol. 28, no. 1, pp. 237–253, June 1998.