WCFQ: an Opportunistic Wireless Scheduler with Statistical Fairness Bounds

Yonghe Liu, Stefan Gruhl, and Edward W. Knightly

***Abstract*—In this paper, we present Wireless Credit-based Fair Queueing (WCFQ), a new scheduler for wireless packet networks with provable statistical short- and long-term fairness guarantees. WCFQ exploits the fact that users contending for the wireless medium will have different "costs" of transmission depending on their current channel condition. For example, in systems with variable coding, a user with a high-quality channel can exploit its low-cost channel and transmit at a higher data rate. Similarly, a user in a CDMA system with a high quality channel can use a lower transmission power. Thus, WCFQ provides a mechanism to exploit inherent variations in channel conditions and select low cost users in order to increase the system's overall performance (e.g., total throughput). However, opportunistic selection of the best user must be balanced with fairness considerations. In WCFQ, we use a credit abstraction and a general "cost function" to address these conflicting objectives. This provides system operators with the flexibility to achieve a range of performance behaviors between perfect fairness of temporal access independent of channel conditions, and purely opportunistic scheduling of the best user without consideration of fairness. To quantify the system's fairness characteristics within this range, we develop an analytical model that provides a statistical fairness bound in terms of the cost function and the statistical properties of the channel. An extensive set of simulations indicate that the scheme is able to achieve significant throughput gains while balancing temporal fairness constraints.**

***Index Terms*—Scheduling, weighted fair queuing, probabilistic fairness guarantee, 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 [1], [2], [3], [4], [5], [6], [7], [8]. In particular, in error-prone wireless links with a binary channel model (0 or 100% link error, as considered in [1], [3], [4], [5], [6] for example), 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 [5] 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, physical proximity to a base station, etc. In general, users can be viewed as having different "costs" to transmit at each particular instance. For example, in a system with variable coding, users with high quality or low-cost channels would be able to transmit at higher data rates. In CDMA systems, low-cost users could reduce their transmission power for a particular throughput level.

Y. Liu and E. Knightly are with the Department of Electrical and Computer Engineering, Rice University. S. Gruhl is with Bell Labs, Lucent Technologies. The authors may be reached via http://www.ece.rice.edu/networks.

Thus, in this case of a continuous channel model, the selection of which user to transmit at a particular instance has important implications on both overall system performance (e.g., total throughput) as well as user fairness properties. Hence, the scheduling and medium access algorithms that select the next user to transmit must be designed to incorporate these effects.

In continuous-channel systems, an alternate view of fairness is more suitable. In particular, the system should provide fair *temporal* access to the medium rather than fair throughput, i.e., ensure that each user is able to access the medium for a (weighted) fair share of time. In the simplified binary channel model, the distinction between temporal and throughput fairness is inconsequential: users can access the channel when they are in an error-free state, and cannot otherwise. However, ensuring temporal rather than throughput fairness has two advantages in continuous channel systems: first, it allows the system to exploit the good channel conditions of high-throughput users without penalty. Second, it provides true "performance isolation" so that a user with a poor channel condition cannot reduce the throughput of other users to arbitrarily low levels while the poor-channel user catches up.

In this paper, we design and analyze Wireless Credit-based Fair Queueing (WCFQ), a scheduler with provable statistical temporal fairness properties over both short- and long-term horizons. Our key technique is to exploit temporal variations in the "cost" of scheduling different users to opportunistically select users with greater throughput potential, while also ensuring that the system's temporal fairness constraints are satisfied.

In particular, we use the credit abstraction to balance a user's cost of accessing a channel with the elapsed time since its prior transmission. As in previous credit-based scheduling schemes [9], users attain credits as they wait to be scheduled. However, rather than select the user with the largest number of credits, the credit counts are compared with the *cost* of selecting a particular user. The cost function reflects the channel quality and the fact that a user with a high quality channel can transmit at a higher throughput or lower resource consumption. Hence, selecting users while they are in low-cost states allows the system to increase its total efficiency in terms of throughput or total power consumption. However, to (statistically) ensure fairness, users will eventually be scheduled by either obtaining a high quality/low cost channel via fluctuations in channel conditions, or instead by obtaining sufficient credits to overcome persistently high channel costs.

By considering a general cost function, we provide system designers with a flexible way to trade off the extent to which total throughput is valued at the expense of fairness and vice versa. For example, at one end of the spectrum, a perfect and deterministically weighted fair schedule can be achieved independent of transmission and power costs. At the other end of the spectrum, the highest throughput or lowest cost user can always be selected without fairness consideration. In between, WCFQ balances this tradeoff with controlled and predictable fairness properties.

To quantify the fairness of the system, we define a probabilistic and time-share fairness index, as opposed to the traditional deterministic and throughput-share fairness index of [10]. We then derive an expression for WCFQ's fairness index as a function of a statistical characterization of the channel and the system's cost function. In this way, we address the random nature of the wireless channel, allow the scheduler to exploit varying channel conditions, and simultaneously ensure that fairness guarantees *statistically* hold for long- and short-term periods.

Finally, we perform an extensive set of simulations to evaluate the performance of WCFQ. To incorporate both mobility and fast fading, we consider a channel model with both slow- and fast-time-scale variations in channel quality (and hence cost). We present a set of scheduling "visualizations" that illustrate WCFQ's temporal operation via depictions of channel quality, accumulated credits, and selected packets as a function of time. Moreover, we show that by using different cost functions, significant gains in throughput can be achieved while simultaneously ensuring different levels of temporal fairness.

The remainder of this paper is organized as follows. In Section II, we describe the channel model and present a stochastic fairness index. In Section III, we present the WCFQ service discipline, analyze its fairness properties, and explore the role of the cost function as it relates to the fairness guarantees. Next, we present the simulation experiments and temporal visualizations in Section IV. Finally, we review related work in Section V and conclude in Section VI.

II. System Model and Stochastic Fairness

In this paper, we consider centralized scheduling for a shared wireless channel that is accessed by multiple users in a time division multiple access manner, i.e., at each time, only one user can transmit over the channel. Furthermore, there is a central scheduler controlling access to the channel. Downlink scheduling is realized by the base station, whereas uplink scheduling uses an additional mechanism such as polling to collect transmission requests from mobile nodes and perform centralized uplink scheduling. We assume the downlink and uplink transmission are separated and do not interfere with each other. Example systems that satisfy this requirement are TDMA schemes like GPRS [11] or CDMA schemes that operate in two distinctive frequency bands.

A. Channel Model

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. The particular mechanism employed by a system depends on the communication standard. For example, in HDR [12] and UMTS-HS-DPA [13], the underlying physical channel uses explicit channel notification so that the scheduler has the best possible knowledge about the channel conditions. In UMTS-DCH [14], there is a logical control channel assigned with every user that allows a coarse estimation of the channel condition. The packet extensions (E)GPRS [11] to GSM-TDMA systems offer various coding schemes to support data transmission over a wide range of channel conditions. These are typically switched on a slower timescale, e.g., based on experienced frame error rates. Regardless, the recently selected coding scheme that determines the 'throughput per RLC-block' can serve as a coarse indicator of the channel condition for the scheduler. In general, the faster and more precisely the channel quality can be predicted, the better the scheduler can incorporate this information into its decision as to which user to schedule next.

To obtain a scheduling algorithm applicable to the above class of standards and systems, we generalize the channel condition into a cost function based on the underlying physical-layer information. The transmission cost reflects the system efficiency due to selecting a particular flow. For example, a user that currently has a poor quality channel will have a high cost to reflect that scheduling that user immediately would require increased transmission power or strong forward error protection that results in lower system utilization. The cost function is a non-negative and non-decreasing function of the channel quality indication. In particular, we denote $U_i(p)$ as the cost of scheduling user i as the p^{th} packet transmission, so that the cost is dynamically updated to reflect changing channel conditions. In our simulation experiments of Section IV, we consider channel conditions that range from 0 to 1, where 0 is the lowest cost or best channel condition.

B. Stochastic Fairness Measure

Our goal is to design a scheduler that can balance the conflicting objectives of achieving high overall system performance and providing weighted fair temporal access to the channel. Here, we provide a formal definition of statistical fairness to quantify this constraint.

The proportional fairness index applied in wire-line networks characterizes the service discrepancy between two flows i and j over any interval (t_1, t_2) during which the two flows are continuously backlogged. Normally, a wire-line scheduler guarantees the proportional fairness index to have a hard upper bound, i.e.,

$$\left|\frac{W_i(t_1,t_2)}{\phi_i} - \frac{W_j(t_1,t_2)}{\phi_j}\right| \le \text{constant}(i,j) \quad (1)$$

where $W_i(t_1, t_2)$ denotes the service in bits that flow i receives during (t_1, t_2), ϕ_i denotes the assigned weight for flow i, and the constant may be a function of the flow indexes i and j.

To design a fair scheduler for wireless networks, we consider two modifications to this index. First, we require a *statistical* fairness index as it provides the scheduler with the flexibility to exploit short term channel variations and select users with better channel conditions. Moreover, it better reflects the randomness inherent in the wireless system's channel conditions.

Second, we consider fairness of flows' channel access times rather than throughputs. The key motivation for this is that in wireless networks, users can transmit at different rates depending on their channel quality. Thus, to normalize throughputs would require allowing the user with the worst channel quality to have a disproportionately large share of channel access time, thereby degrading overall system throughput. Thus, temporal fairness attains the "isolation" property in which a user entering a region with a persistently poor channel condition has a controlled and predictable effect on other users' throughputs, whereas to obtain throughput fairness, the bad-channel user could reduce the throughput of other users to an unpredictable level.

Hence, to make the distinction between temporal and throughput fairness, we define $\alpha(t_1, t_2)$ as the service in *time* that flow i receives during (t_1, t_2). Moreover, to relax the former fairness guarantee to be a statistical one, we define a statistical time-access fairness index as:

$$Pr\left(|\frac{\alpha_i(t_1,t_2)}{\phi_i} - \frac{\alpha_j(t_1,t_2)}{\phi_j}| \geq x\right) \leq f(i,j,x) \qquad (2)$$

This index reflects our WCFQ design objective: if a user enters a region of poor channel quality, we statistically maintain its temporal share of the channel, but do not attempt to normalize all flows' cumulative throughputs. Instead, we will show that with WCFQ, by opportunistically selecting users that are now transmitting on higher-rate channels, the system can attain significantly higher throughput while maintaining statistical temporal fairness and performance isolation.

III. Stochastic Wireless Scheduling

In this section, we first present a new wireless scheduling discipline termed Wireless Credit-based Fair Queuing (WCFQ). We then derive an expression for WCFQ's statistical fairness characteristics that allow WCFQ systems to provide statistically-fair channel access guarantees. Finally, we explore the role of the cost function in enabling network operators to trade stricter fairness for higher throughput while still maintaining quantifiable fairness characteristics.

A. *Wireless Credit-based Fair Queuing - WCFQ*

Our goal in designing a wireless service discipline is to provide the rigorous fairness guarantees typically associated with wire-line networks, while simultaneously employing opportunistic scheduling strategies to increase the total system throughput by selecting users with high-quality channels when possible. To achieve this, we incorporate users' channel conditions into the scheduling decision while also balancing fairness constraints via the abstraction of *credits*.

In wire-line networks, Credit Based Fair Queueing (CBFQ) was introduced to achieve the same proportional throughput fairness as WFQ in a more computationally efficient way [9]. The technique is to utilize a single status value per *flow*, termed a credit, and thereby avoid the per-*packet* tags of WFQ and its variants [15], [5], [16]. In CBFQ, flows accumulate credits when they are not scheduled whereas credits are decremented when the flow is scheduled. By assigning credits according to the number of backlogged flows, relative weights, etc., the CBFQ scheduling decision is simply to select the packet from the head of the queue with the smallest value of a specially designed function of the credits, weights, etc.

While WCFQ shares the computational advantages of CBFQ, our primary use of a credit based scheme is to incorporate both the channel condition and fairness into the scheduling decision. In particular, we define Wireless Credit-based Fair Queuing as follows. Consider N users accessing a shared channel with user i have weight ϕ_i such that the weight represents user i's targeted temporal share of the channel. Furthermore, let p denote the index of the p^{th} packet transmitted over the channel. For ease of notation, we assume WCFQ's scheduling decision occurs at the end of a packet transmission epoch such that p can also be considered to be a time index for all dynamic parameters including channel conditions.

TABLE I
Summary of Notation

Term	Definition
p	index of the packet in service
i	session index
ϕ_i	weight of flow i
f_p	flow that packet p belongs to
$L_i(p)$	HOL packet length for flow i at p
$K_i(p)$	credit counter of flow i at p
$B(p)$	the set of backlogged flows at p
$U_i(p)$	estimated cost for user i to transmit the p^{th} packet
$E_i(p)$	channel condition for user i at p

Let f_p denote the flow that transmits the p^{th} packet, and let $L_i(p)$ denote the actual packet transmission time of the head of line (HOL) packet of flow i over the wireless channel when the p^{th} packet's transmission ends. This time is a function of the underlying transmission scheme such as the coding scheme, modulation scheme, and spreading factor in a CDMA network. Finally, denote $B(p)$ as the set of backlogged flows and $K_i(p)$ as the credit value for flow i at time p. Notation is summarized in Table I.

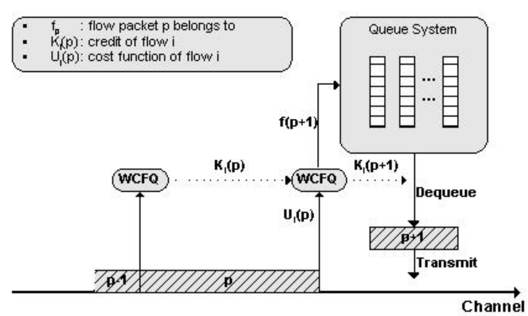


Fig. 1. Conceptual Operation of WCFQ

We now define the WCFQ scheduling algorithm by describing how credits are initially set, how they are used to make a scheduling decision, how they are increased when a flow is not scheduled, and how a credit is decreased when a flow is scheduled. Figure 1 illustrates the conceptual operation of WCFQ and the algorithm is described as follows.

When flow i enters the system or becomes unbacklogged, K_i is set to zero. Once the server finishes transmitting packet p, WCFQ will select flow f_{p+1} from the backlogged set to decide the $(p+1)^{th}$ packet to be served. The key difference between WCFQ and CBFQ is this session selection criteria, as with WCFQ we also incorporate the transmission cost in order to balance a flow's cumulated credits with the current channel condition, thereby balancing system fairness and total throughput. In particular, if the current transmission cost of a user is high, the scheduler may postpone its packet transmission despite its cumulated credits. However, this access delay is limited as the user will eventually obtain sufficient credits (due to a lack of being scheduled) to outweigh a poor channel condition. This characteristic is reflected in the following basis for scheduling decisions, where we apply the cost function in a sum with the credits. That is, the packet selection criteria is to select the HOL packet from the flow that satisfies:

$$f_{p+1} = arg\min_{i\in B(p)} \frac{L_i(p)-K_i(p)+U_i(p)}{\phi_i} \qquad \text{(a)}$$

Following this scheduling decision, all active sessions update their credits as described via pseudo-code in Table II.

TABLE II
PSEUDO-CODE FOR CREDIT UPDATE

for ($i = 1;\ i \le N;\ i++$)
  if ($i \in B(p+1)$ && $i \neq f_{p+1}$)
    $K_i(p+1)$
    $= K_i(p) + \max(\frac{L_{f_{p+1}}(p)-K_{f_{p+1}}(p)}{\phi_{f_{p+1}}}, 0)\phi_i$ (b)
  elseif ($i \notin B(p+1)$)
    $K_i(p+1) = 0$ (c)
  end
end

if ($f_{p+1} \in B(p+1)$)
  $K_{f_{p+1}}(p+1) = \max(0, K_{f_{p+1}}(p) - L_{f_{p+1}}(p))$ (d)
else
  $K_{f_{p+1}}(p+1) = 0$ (e)
end

This procedure is reasoned as follows. As in previous WFQ algorithms, un-backlogged flows are idle in the system and their credits are always set to zero (c, e). This is reasonable as fairness is only meaningful for backlogged flows. The currently scheduled flow decreases its credit count, but unlike CBFQ not automatically to 0. If the flows's credit ($K_{f_{p+1}}(p)$) is larger than its current HOL packet length ($L_{f_{p+1}}(p)$), it is only decreased by the length of the selected packet (d). This allows backlogged flows to cumulate credits over a longer period so that a flow with persistently poor channel conditions does not instantly lose all of its credits when scheduled. If on the other hand a session is scheduled before its credits reach the value of the HOL packet, we reset its credit to zero. Both cases are combined in the *max* function (d). Finally, all non-scheduled flows increase their credit counts in a weighted fair manner. If the selected flow's credit is larger than its current HOL packet length, we set this increase to be zero. Otherwise we proceed as in CBFQ and add the relative share to the credit, again combined with previous rule using a *max* function (b).

We make two observations about WCFQ. First, as it is inherited from CBFQ, it does not require per packet tags. Its operations consist of a flow selection and an update step. For a system with n active flows, the former operation consists of the computation of $U(p) + K(p)$ per-flow operations at cost $O(n)$ and the selection of the flow with the lowest value at cost $O(log(n))$. The update requires one addition and one division per flow again at cost $O(n)$. Thus, WCFQ is computationally feasible in many wireless systems that have a moderate number of flows per base station.

Second, note that the design of the decision function (a) that combines credits and channel functions requires that the unit of transmission cost match the unit of packet length. Selection of the transmission cost function balances fairness vs. the total system throughput as discussed further in Section III-C.

B. Fairness Analysis

With WCFQ, we target to fairly allocate the shared channel's time slots to sessions according to their weight. To characterize the statistical fairness property of WCFQ, we derive a statistical bound on the weighted difference in allocated time slots for any flows i and j that are continuously backlogged in (t_1, t_2):

$$|\frac{\alpha_i(t_1,t_2)}{\phi_i} - \frac{\alpha_j(t_1,t_2)}{\phi_j}|$$

The key idea to derive this statistical fairness property is to characterize the fairness bound as a function of the transmission cost. Since the transmission cost is a function of the randomly varying channel conditions, the channel conditions, transmission costs, and hence fairness are all characterized statistically.

Before analyzing the fairness of WCFQ, we describe the difference between the credit counter in WCFQ and CBFQ, as it is crucial to understand the statistical fairness property of WCFQ. CBFQ always limits a flow's credit to be between zero and its maximum packet length. This allows CBFQ to provide a hard bound on the weighted fairness index. However, WCFQ allows a flow to accumulate credits without any hard limit. For example, WCFQ allows the credit counter to become much larger due to a user having poor channel conditions for several continuous scheduling intervals. This distinction provides the flexibility to dynamically adjust scheduling decisions to channel conditions.

The analysis is organized as follows. We first show the credit counter for a flow is upper and lower bounded in *Lemma 1*, where the upper bound is related to the transmission cost function. Then in *Lemma 2*, the received service of a flow during interval (t_1, t_2) is expressed as a function of the credit values at t_1 and t_2. We then statistically bound the proportional service discrepancy of any two flows in *Theorem 1*.

Lemma 1: For any two flows i and j, $\exists p' \leq p$, such that their accumulated credits $K_i(p+1)$ and $K_j(p+1)$ are bounded by

$$0 \leq K_i(p+1) \leq L_i^{\max} + U_i(p') \quad (3)$$
$$0 \leq K_j(p+1) \leq L_j^{\max} + U_j(p') \quad (4)$$

where $L_i^{\max} = max(L_i)$.

To prove *Lemma 1*, we first show that the inequality holds for an individual flow, and then prove that the inequality is synchronized for any two flows.

Proposition 1: For any flow i , $\exists\, p' \leq p$, such that its accumulated credit count $K_i(p+1)$ is bounded by

$$0 \leq K_i(p+1) \leq L_i^{\max} + U_i(p') \quad (5)$$

Proof: Credits are either increased according to operation (b) in the scheduler specification of Table II, or decreased to no lower than 0 according to operation (d). Thus $0 \leq K_i(p+1)$. For the right part of the inequality, the proof is separated into two cases according to whether or not flow i is scheduled after packet p's transmission. In each case, there are two sub-cases regarding how the credit is updated.

Case I: $i \neq f_{p+1}$

If $K_{f_{p+1}}(p) - L_{f_{p+1}}(p) < 0$, then from (a) we have

$$\frac{L_{f_{p+1}}(p) - K_{f_{p+1}}(p) + U_{f_{p+1}}(p)}{\phi_f} \leq \frac{L_i(p) - K_i(p) + U_i(p)}{\phi_i}$$
$$K_i(p) + \frac{L_{f_{p+1}}(p) - K_{f_{p+1}}(p) + U_{f_{p+1}}(p)}{\phi_{f_{p+1}}}\phi_i \leq L_i + U_i(p)$$

Considering the credit update rule (b), we have

$$\begin{aligned} K_i(p+1) &\leq L_i(p) + U_i(p) - \frac{U_{f_{p+1}}(p)}{\phi_{f_{p+1}}}\phi_i \\ &\leq L_i^{max} + U_i(p) \end{aligned}$$

The last step holds because we have limited our transmission cost to be non-negative, i.e., $U_{f_{p+1}}(p) \geq 0$.

If $K_{f_{p+1}}(p) - L_{f_{p+1}}(p) \geq 0$, then from step (b), $K_i(p+1) = K_i(p)$. For $K_i(p)$, the proof is the same as for $K_i(p+1)$. This backward process can be done recursively until a packet p' when

- flow i is not scheduled and $K_{f_{p'+1}}(p') - L_{f_{p'+1}}(p') < 0$, or
- flow i is not backlogged when $K_i(p') = 0$, or
- flow i is scheduled which is dealt in *Case II*.

Case II: $i = f_{p+1}$,

If $K_{f_{p+1}}(p) - L_{f_{p+1}}(p) < 0$, $K_{f_{p+1}}(p+1) = 0$ from step (c), and Lemma 1 holds.

If $K_{f_{p+1}}(p) - L_{f_{p+1}}(p) \geq 0$, then from step (d) we have

$$K_{f_{p+1}}(p+1) = K_{f_{p+1}}(p) - L_{f_{p+1}}(p) \leq K_{f_{p+1}}(p). \quad (6)$$

Analogous to *Case I*, $K_i(p)$ can be bounded recursively until a packet p' when

- flow i is scheduled and $K_{f_{p'+1}}(p') - L_{f_{p'+1}}(p') < 0$, or
- flow i is not backlogged when $K_i(p') = 0$, or
- flow i is not scheduled which is dealt in *Case I*.

From *Cases I* and *II*, we have $\exists\, p',\ s.t.$

$$0 \leq K_i(p+1) \leq L_i^{max} + U_i(p')$$

■

Using *Proposition* 1, we can prove *Lemma 1* as follows.

Proof: Notice that for flow i, a p' is selected when either of the following two conditions is satisfied

- $K_{f_{p'+1}}(p') - L_{f_{p'+1}}(p') < 0$, or
- flow i is not backlogged.

The first case is common to every flow. Thus, if this is the situation for both flow i and j, *Lemma 1* holds. If the second situation is satisfied first for a flow, e.g., i, when packet $p'' \leq p$ is transmitting, we must have $K_i(p+1) \leq K_i(p'') = 0$ according to *Proposition 1* and hence $K_i(p+1) = 0$. Therefore, for any packet $p' \leq p$, we have $K_i(p+1) = 0 \leq L_i^{max} + U_i(p')$. Considering both cases, *Lemma 1* holds. ■

Notice that $K_i(p)$ does not change while packet p is being transmitted. Define $K_i(t) = K_i(p)$ if packet p is being transmitted at time t.

Lemma 2: Assume packets $(1, \cdots, p, \cdots, N),\ N \geq 1$ start transmitting during time period (t_1, t_2). For any flow i continuously backlogged during (t_1, t_2), its received service $\alpha_i(t_1, t_2)$ during (t_1, t_2) can be expressed as:

$$\begin{aligned} K_i(t_1) + \sum_{p=0}^{N-1} \max\left(\frac{L_{f_p}(p) - K_{f_p}(p)}{\phi_{f_p}}, 0\right)\phi_i & \\ = \quad \alpha_i(t_1, t_2) + K_i(t_2) \end{aligned}$$

Proof: If no packet of flow i is served during (t_1, t_2), $\alpha_i(t_1, t_2) = 0$. According to line (b) of the WCFQ scheduling algorithm, equality holds. If some packet $(p+1) \in 1, \cdots, N$ of flow i is served during (t_1, t_2), there are two cases:

Case I: if $K_{f_p}(p) - L_{f_p}(p) \leq 0$,[1] then, according to step (b) of WCFQ, $K_{f_p}(p+1) = 0$. On the other hand,

$$\begin{aligned} & K_{f_p}(p) + \max(\frac{L_{f_p} - K_{f_p}(p)}{\phi_{f_p}}, 0)\phi_{f_p} \\ = \quad & K_{f_p}(p) + \frac{L_{f_p} - K_{f_p}(t)}{\phi_{f_p}}\phi_{f_p} \\ = \quad & L_{f_p} + K_{f_p}(p+1) \end{aligned} \quad (7)$$

Case II: if $K_{f_p}(p) - L_{f_p}(p) > 0$, then according to step (b) of the scheduling algorithm, $K_{f_p}(p+1) = K_{f_p}(p) - L_{f_p}(p)$. On the other hand,

$$\begin{aligned} & K_{f_p}(p) + \max(\frac{L_{f_p} - K_{f_p}(p)}{\phi_{f_p}}, 0)\phi_{f_p} \\ = \quad & K_{f_p}(p) + 0 \\ = \quad & L_{f_p} + K_{f_p}(p+1) \end{aligned} \quad (8)$$

Thus, from Equations (7) and (8), *Lemma 2* holds. ■

Theorem 1: For any two flows i and j continuously backlogged over any interval (t_1, t_2), we have the following fairness guarantee: $\exists\ p'_i,\ p'_j$, s.t.

$$\begin{aligned} & Pr\left(|\frac{\alpha_i(t_1, t_2)}{\phi_i} - \frac{\alpha_j(t_1, t_2)}{\phi_j}| \geq \frac{L_i^{\max} + x}{\phi_i} + \frac{L_j^{\max} + x}{\phi_j}\right) \\ \leq \quad & Pr\left(\frac{U_i(p'_i)}{\phi_i} + \frac{U_j(p'_j)}{\phi_j} \geq \frac{x}{\phi_i} + \frac{x}{\phi_j}\right) \end{aligned}$$

[1] Note that to schedule packet $p+1$, the decision is being made when packet p is being transmitted.

Proof: If no packets start transmitting during (t_1, t_2), *Theorem 1* holds. Otherwise, from *Lemma 2*, we have

$$\begin{aligned}&\alpha_i(t_1,t_2)\\ &= K_i(t_1)+\sum_{p=0}^{N-1}\max\left(\frac{L_{f_p}(p)-K_{f_p}(p)}{\phi_{f_p}},0\right)\phi_i-K_i(t_2)\end{aligned}$$

Therefore, the service discrepancy is given by

$$\begin{aligned}&|\frac{\alpha_i(t_1,t_2)}{\phi_i}-\frac{\alpha_j(t_1,t_2)}{\phi_j}|\\ &= |\frac{K_i(t_1)-K_i(t_2)}{\phi_i}-\frac{K_j(t_1)-K_j(t_2)}{\phi_j}|\\ &\le |\frac{K_i(t_1)-K_i(t_2)}{\phi_i}|+|\frac{K_j(t_1)-K_j(t_2)}{\phi_j}|\\ &\le |\frac{\max(K_i(t_1),K_i(t_2))}{\phi_i}|+|\frac{\max(K_j(t_1),K_j(t_2))}{\phi_j}|\end{aligned}$$

Applying *Lemma 1*, we have that $\exists\, p_i',\ p_j'$ such that

$$\begin{aligned}&|\frac{\alpha_i(t_1,t_2)}{\phi_i}-\frac{\alpha_j(t_1,t_2)}{\phi_j}|\\ &\le \frac{L_i^{\max}+U_i(p_i')}{\phi_i}+\frac{L_j^{\max}+U_j(p_j')}{\phi_j}\end{aligned} \tag{9}$$

Therefore, by relaxing the service discrepancy using Equation (9), we have

$$\begin{aligned}&Pr\left(|\frac{\alpha_i(t_1,t_2)}{\phi_i}-\frac{\alpha_j(t_1,t_2)}{\phi_j}|\ge\frac{L_i^{\max}+x}{\phi_i}+\frac{L_j^{\max}+x}{\phi_j}\right)\\ &\le Pr\left(\frac{U_i(p_i')}{\phi_i}+\frac{U_j(p_j')}{\phi_j}\ge\frac{x}{\phi_i}+\frac{x}{\phi_j}\right)\end{aligned}$$

■

If we further assume that the transmission cost is i.i.d. in each time slot, we have the following corollary. Let the random variable U_i denote the transmission cost for flow i.

Corollary 1: For any two flows i and j with cost function U_i and U_j continuously backlogged over any interval (t_1, t_2), we have the following fairness guarantee:

$$\begin{aligned}&Pr\left(|\frac{\alpha_i(t_1,t_2)}{\phi_i}-\frac{\alpha_j(t_1,t_2)}{\phi_j}|\ge\frac{L_i^{\max}+x}{\phi_i}+\frac{L_j^{\max}+x}{\phi_j}\right)\\ &\qquad\le Pr\left(\frac{U_i}{\phi_i}+\frac{U_j}{\phi_j}\ge\frac{x}{\phi_i}+\frac{x}{\phi_j}\right)\end{aligned}$$

Proof: Follows directly from *Theorem 1* with the i.i.d. assumption. ■

Corollary 1 seems loose when $(t_2 - t_1)$ is small, especially when x is large. A simple way to tighten it is as follows. Notice that

$$|\frac{\alpha_i(t_1,t_2)}{\phi_i}-\frac{\alpha_j(t_1,t_2)}{\phi_j}|\le\max\left(\frac{\alpha_i(t_1,t_2)}{\phi_i},\frac{\alpha_j(t_1,t_2)}{\phi_j}\right)$$

since $\alpha_i(t_1, t_2)$ is nonnegative for any i. Therefore,

$$Pr\left(|\frac{\alpha_i(t_1,t_2)}{\phi_i}-\frac{\alpha_j(t_1,t_2)}{\phi_j}|\ge\frac{L_i^{\max}+x}{\phi_i}+\frac{L_j^{\max}+x}{\phi_j}\right)=0$$

when

$$\begin{aligned}&\max\left(\frac{\alpha_i(t_1,t_2)}{\phi_i},\frac{\alpha_j(t_1,t_2)}{\phi_j}\right)\\ &\qquad\le\left(\frac{L_i^{\max}+x}{\phi_i}+\frac{L_j^{\max}+x}{\phi_j}\right).\end{aligned} \tag{10}$$

Therefore, *Corollary 1* can be rewritten as follows.

$$\begin{aligned}&Pr\left(|\frac{\alpha_i(t_1,t_2)}{\phi_i}-\frac{\alpha_j(t_1,t_2)}{\phi_j}|\ge\frac{L_i^{\max}+x}{\phi_i}+\frac{L_j^{\max}+x}{\phi_j}\right)\\ &\le\min\left(Pr\left(\frac{U_i}{\phi_i}+\frac{U_j}{\phi_j}\ge\frac{x}{\phi_i}+\frac{x}{\phi_j}\right),1-I\left(T\right)\right),\end{aligned} \tag{11}$$

Where T is denoted by (10) and I is the indication function.

Equation (11) depicts that when (t_2-t_1) increases, the probabilistic bound regulates the behavior of the scheduler more and more. However, different from a long-term fairness definition on the infinite horizon, it is defined on any time interval and thus much “tighter”.

C. Fairness Guarantees through Cost Function Design

Having derived the fairness bound in terms of the cost function distribution, one of the key problems to address is how network operators should select a proper transmission cost function. In other words, given the desired statistical fairness requirements and channel distribution, what transmission cost function should we define?

As described in Section II, different systems will have different cost functions to translate information from the physical layer channel conditions. For example, a CDMA system may use power consumption in the cost function while a time division system may use the coding rate. Given the distribution of the channel characteristics, we now need to specify the mapping from the physical information to the transmission cost to achieve the desired fairness property. Next, we give an example of how the transmission cost function can be derived: a similar methodology can be performed for different cases.

Denote E_i as a random variable characterizing the channel condition of user i. In order to find the proper transmission cost, the operator will first specify the desired level of fairness. For example, for two sessions i and j, it may require

$$\begin{aligned}&Pr\left(|\frac{\alpha_i(t_1,t_2)}{\phi_i}-\frac{\alpha_j(t_1,t_2)}{\phi_j}|\ge\frac{L_i^{\max}+x}{\phi_i}+\frac{L_j^{\max}+x}{\phi_j}\right)\\ &\le g(x)\end{aligned}$$

to be satisfied for some function $g(\cdot)$. With *Corollary 1*, this is achieved if the transmission cost function satisfies

$$Pr(\frac{U_i}{\phi_i}+\frac{U_j}{\phi_j}\ge\frac{x}{\phi_i}+\frac{x}{\phi_j})\le g(x). \tag{12}$$

To simplify, consider the bound

$$\begin{aligned}&Pr(\frac{U_i}{\phi_i}+\frac{U_j}{\phi_j}\le\frac{x}{\phi_i}+\frac{x}{\phi_j})\\ &\qquad\ge Pr(\frac{U_i}{\phi_i}\le\frac{x}{\phi_i})*Pr(\frac{U_j}{\phi_j}\le\frac{x}{\phi_j})\end{aligned} \tag{13}$$

Therefore, if U_i and U_j have the same distribution, we can further simplify the requirement to be

$$Pr(U_i \leq x) \geq \sqrt{1-g(x)}. \quad (14)$$

Thus, based on the distribution of the channel condition, we can find a function $f(\cdot)$ such that by defining $U_i = f(E_i)$ Inequality (14) is satisfied.

To continue with the example, consider a probabilistic fairness requirement that decreases rapidly towards 0 for increasing x as given by:

$$g(x) = 2\exp(-\frac{x}{\beta}) - \exp(-\frac{2x}{\beta}), x \geq 0. \quad (15)$$

Notice that β is a tunable parameter to tradeoff between fairness and system gain (a tradeoff further explored in Section IV). Assume E_i has a uniform distribution in $[0, 1]$. Then we can set

$$U_i = -\beta\, log(1 - E_i). \quad (16)$$

The above example illustrates that by defining different transmission costs, the system can obtain stronger or weaker fairness guarantees. Intuitively, with larger possible values given to the transmission cost, the channel quality weighs more heavily into WCFQ's packet selection decision. Consequently, a greater total system throughput will be achieved at the expense of a looser fairness constraint.

Most of the time, an analytical model for the users' channel condition is not readily available. However, an operator may have field measurements and an approximation to this distribution. With the desired fairness objective, the operator can design a map between the channel condition and cost function. For example, Figure 2 depicts the distribution of the channel condition (e.g., E) for a typical user and how it can be mapped to the desired cost function derived from the fairness requirement.

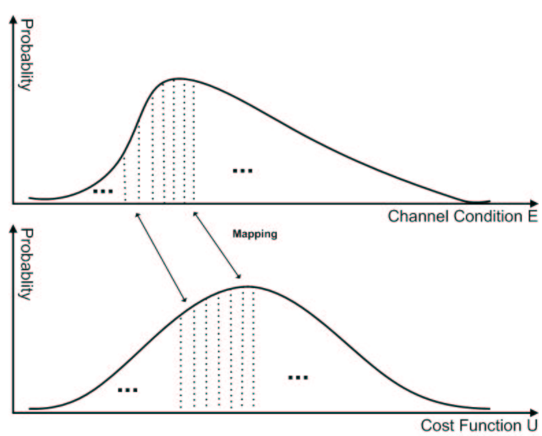


Fig. 2. Mapping Channel Condition to Cost Function

Furthermore, with certain selections of cost functions, WCFQ degenerates into several special wireless and wire-line schedulers. These include the scheduling algorithms based simply on the binary channel model and the wire-line weighted fair scheduler CBFQ. In this sense, WCFQ can be viewed as a generalized version of wireless schedulers, which stems from its flexible statistical fairness guarantee.

For example, the cost function can be assigned to be zero or infinite according to whether the channel quality is above or below a certain threshold. WCFQ hence degrades into a scheduler based on the binary model assumption.

One extreme is that the transmission cost is set to zero for any flow. In this case, the scheduler degenerates to CBFQ and *Theorem 1* becomes a hard fairness guarantee. More precisely, for any $x, x \geq 0$, the following equation holds:

$$\begin{aligned} Pr\left(|\frac{\alpha_i(t_1,t_2)}{\phi_i} - \frac{\alpha_j(t_1,t_2)}{\phi_j}| \geq \frac{L_i^{\max}+x}{\phi_i} + \frac{L_j^{\max}+x}{\phi_j}\right) & \\ \leq \quad Pr\left(\frac{U_i}{\phi_i} + \frac{U_j}{\phi_j} \geq \frac{x}{\phi_i} + \frac{x}{\phi_j}\right) & \\ \leq \quad Pr\left(0 \geq \frac{x}{\phi_i} + \frac{x}{\phi_j}\right) & \\ = \quad 0. & \end{aligned}$$

Therefore, for the special case of $x = 0$, we have that

$$Pr\left(|\frac{\alpha_i(t_1,t_2)}{\phi_i} - \frac{\alpha_j(t_1,t_2)}{\phi_j}| \geq \frac{L_i^{\max}}{\phi_i} + \frac{L_j^{\max}}{\phi_j}\right) = 0$$

which is indeed the traditional "deterministic" fairness index associated with wire-line schedulers.

The alternate extreme is to assign transmission cost functions to generate infinite discrepancy under different channel conditions. In such cases, the channel cost will dominate any accumulated credit, thus making WCFQ simply select the user with the best channel without regard to fairness.

IV. Simulation Experiments

In this section, we present an extensive set of simulation experiments to evaluate the performance of WCFQ. We explore the role of the cost function as a flexible mechanism to tradeoff between the astringency of the fairness guarantee and the total system throughput. We also consider two extreme cases of WCFQ cost functions as baselines for comparison. The first is a cost function that is zero, independent of the channel condition. As described in Section III, in this case WCFQ degenerates to credit based fair queueing (CBFQ). This scenario can be considered as achieving perfect fairness of temporal access by alternating service among flows, independent of their channel condition. Second, we consider a scheme in which the cost function gives a very heavy weight to the transmission cost and simply selects the flow with the best channel condition, independent of the flow's prior relative share of temporal access. This scheme, which we refer to as *best channel condition*, ignores fairness and maximizes the total system throughput by always selecting the best possible user. In both cases, we explore fairness over both short- and long-term horizons.

A. System Model

To explore the role of the channel conditions on system throughput and per-user fairness, we consider the following channel model in which channel condition values range between 0 (good) and 1 (bad). We consider a number of scenarios. In the first set of experiments, we consider a channel model characterized 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

$$E_i(t) = 0.5 + d\cos(2\pi f_{i,t}t + \theta_i) + z_i(t) \tag{17}$$

where $\theta_1, \theta_2, \cdots$ are independent and uniformly distributed in $[0, 2\pi]$ giving the channel conditions statistically independent phases. Moreover, we consider the frequency of the sinusoid to also be a random process such that $f_{i,t}$ is a Gaussian moving average process.

The sinusoidal term represents the long time scale effects of mobility for different mobility speeds and channel time scales $1/f_i$. Since this term is within $0.5 \pm d$, d represents the range of the channel effects due to mobility. The additive noise $z(t)$ represents a model of the effects of Rayleigh and Shadow fading via the conservative assumption of additive white uniform noise in the range $[-w, w]$. For most examples, the range of this fading effect is ± 0.2, but it is varied for other experiments. This simple model allows us to study the influence of the experienced channel on both short- and long-term fairness.

For other experiments, we consider channel models as follows. Here, we have two static users, user 1 with a consistently better channel given by $c_1(t) = 0.2 + z_1(t)$, and user 2 with a consistently worse channel given by $c_2(t) = 0.8 + z_1(t)$. Two mobile users move linearly within $[0.2, 0.8]$ and have channel models $c_3(t) = 0.2 + kt + z_3(t)$ and $c_4(t) = 0.8 - kt + z_4(t)$ with the slope k computed so that the simulation ends when user 3 reaches mean channel condition 0.8 and user 4 reaches mean channel condition 0.2. The goal of this model is to address scenarios with some users having consistently bad channels, others consistently good, and others moving between locations of different channel conditions. In all cases, the mean channel condition averaged over all users is 0.5, and the mobility variation parameter is d and the fading variation parameter is w.

Furthermore, we consider a channel in which all flows experience a statistically similar frame error ratio (FER) over time. This is justified, as in our considered wireless systems, channel adaptation allows for predictable FER, which allows us to statistically ignore RLC retransmissions. By ignoring retransmission, our performance analysis relates to the scheduled packets rather than successfully received packets.

For simplicity, we consider a traffic model with all flows continuously backlogged. With the offered load remaining constant, the achieved fairness is entirely related to the scheduling process and channel conditions without any variation due to traffic fluctuations. Moreover, the packet size is fixed to one RLC block fitting exactly one time slot.

Finally, we consider different cost functions as described in the design example of Section III-C. To analyze a range of cost functions of channel conditions E_i, we consider the cost function parameterized by $\beta > 0$ such that $U_i = -\beta \log(1 - E_i)$. Thus, by varying β in simulations, we investigate the tradeoff between the extreme case of perfect CBFQ fairness ($\beta = 0$) and best-channel-condition scheduling ($\beta = \infty$).

B. Visualization of the Scheduling Process

Figure 3 depicts a temporal visualization of the WCFQ scheduling process. As described in Section III, the scheduling decision in each time-slot is based on the flow with the minimum channel cost function (with smaller values indicating low cost or high quality) minus accumulated credit. To illustrate the relationship between the channel condition and credits, the figure depicts the channel condition inverted as the cost function so that the scheduler can be viewed as selecting the packet with the maximum of the difference between the two values (accumulated credits and channel condition).

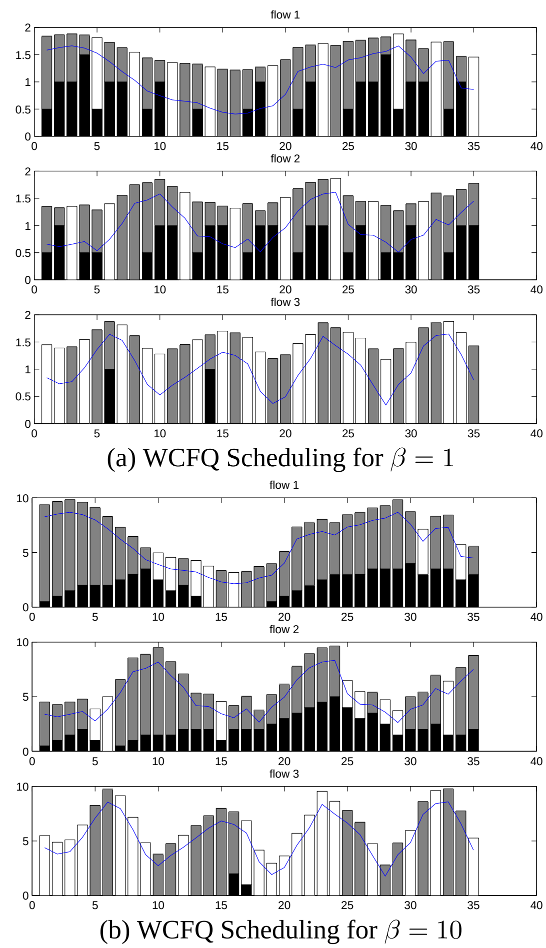


Fig. 3. Visualization of Scheduling Process

Each subgraph shows three bar-graphs, one for each flow. Each bar consists of two components: the lower solid bar represents the accumulated credits whereas the upper bar the cost function of the channel condition. When the bar is colored grey the flow was not selected by the scheduler during this time slot, whereas a white bar indicates that this flow was scheduled.

In addition to the bar presentation, the solid line depicts the channel condition scaled to match the range of the bars. To aid with visualization, the effect of $z_i(t)$ on the channel condition is not depicted. Figure 3(a) shows the system's behavior for $\beta = 1$ whereas Figure 3(b) for $\beta = 10$, to illustrate the role of a factor of 10 in the cost function on scheduling. The flow weights are given by $\phi_1 = \phi_2$ and $\phi_3 = 2\phi_1$ so that flow 3 statistically obtains twice the temporal share of the channel as flows 1 and 2.

Figure 3(a) illustrates that a smaller β (less emphasized cost function) results in WCFQ scheduling the flows more evenly, thereby achieving more stringent fairness over small time scales. Alternatively, Figure 3(b) illustrates that for larger β, flows are not scheduled as periodically so that the overall experienced access ratios are $7/8/20$ rather than $9/8/18$ for $\beta = 1$. The assigned weights of $1:1:2$ are statistically reflected in both examples, and the relaxed fairness constraint of $\beta = 10$ results in a total system throughput gain of approximately 15%.

C. System Throughput

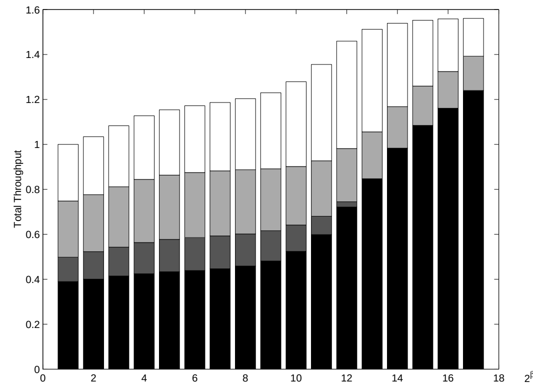
Here, we explore the effect of WCFQ scheduling on total system throughput, where the throughput at transmission epoch p is given by $1 - E_i(p)$. The channel conditions are an important factor, as more widely varying channel conditions provide the scheduler with increased opportunities to select high throughput users. We consider the aforementioned four flow scenario with various values of w and d.

First we study system throughput improvements for various β, the parameter modulating the cost function. Figure 4(a) shows the relative throughput gains whereas Figure 4(b) depicts the relative number of access time slots per flow during a simulation run of 10000 time slots.

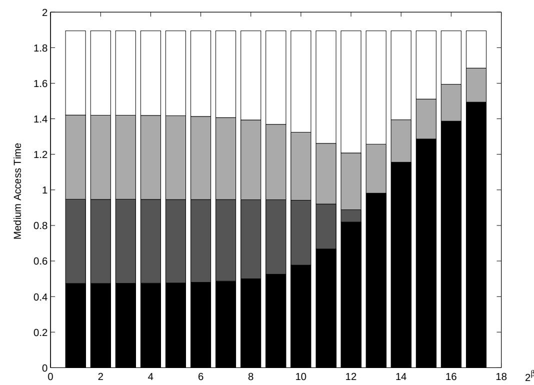
The parameter for short term channel variation is $w = 0.4$ and the parameter for mobility is $d = 0.3$. The leftmost bar represents absolute fairness implemented with CBFQ ($\beta = 0$), achieving the normalized reference system gain of 1. The rightmost bar depicts the best-channel-condition scheduling policy, i.e., always scheduling the flow with the least-cost channel without consideration of fairness ($\beta = \infty$). The intermediate bars show WCFQ scheduling with β ranging exponentially with the values $2^{[0-18]}$.

We make three observations about the figures. First, for the baseline case of CBFQ and $\beta = 0$, the flows achieve identical temporal shares of the channel (2500 time slots), but do not achieve identical throughput due to their different average channel conditions. For example, the black box at the bottom represents the throughput of flow 1, and illustrates that the static user with the constant best channel condition obtains the highest throughput, whereas its neighboring box is flow 2 with the worst channel condition. Second, the asymmetrical share between the two mobile users depicted in the upper part of the bars stems from the fact that the wireless scheduler is more opportunistic and for large cost functions allows a longer time in the future to compensate. Mobile user 3 with the decreasing channel therefore suffers from a reduced share than the uppermost user with the increasing conditions. Third, while increasing β enables higher total throughput, excessive weighting of the channel condition in the packet selection criteria (e.g., bar 14, $\beta > 5000$) has starved flow 2. Regardless, between these extremes, a wide range of β's (and hence cost functions) yield an effective tradeoff between throughput gains and fairness.

Next we study the effects of the channel parameters on system throughput. We first vary w, the range of the uncorrelated uniform channel variations, between 0.1 and 0.4, while keeping $d = 0.3$. One line in Figure 5(a) depicts total system throughput as in the total bar length in Figure 4(a). It is intuitively clear that larger channel variations provide the opportunity for larger throughput gains. Note that the maximum



(a) Normalized Throughput Gains



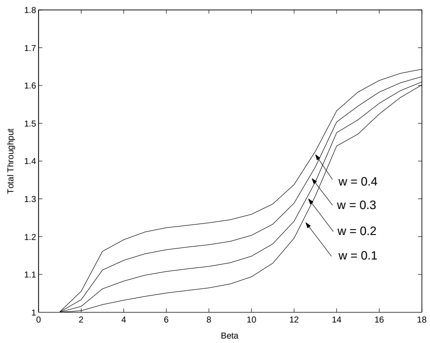
(b) Channel Access

Fig. 4. Efficiency Study for Different Cost Functions: $\beta = 0, 2^{[0-15]}, \infty$

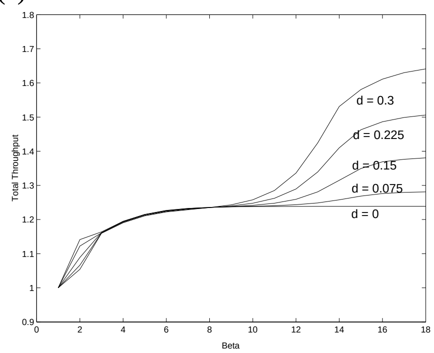
gain using best-channel-condition scheduling increases for the larger channel variations of 0.4 only marginally from 60% to less than 65%. The non-linear shape of the overall gain near $\beta = 2^{10}$ through $\beta = 2^{14}$ is partly due to the exponential cost function that reaches saturation from $\beta > 2^{14}$.

Figure 5(b) depicts simulations with $w = 0.4$ and decreased distance between channels via a reduction in the distance spread value d to 0, i.e., all flows have a constant and statistically identical mean channel, 0.075, 0.15, 0.225, and finally 0.3 as used in the previous example. With smaller d, all channel conditions remain close to of 0.5 and there is less room for opportunistic scheduling. Thus, all values of d obtain similar behavior for small β, whereas with larger values of β, higher gains can be achieved.

We remark that under channels with different mean, the gain is a function of the session duration (or in this context, simulation duration), as the bad channel user can eventually catch up and decrease the gain. This yields the same effect as when β is small and the fairness bound prevents the good channel user from capturing the channel. For sessions of longer duration, the same throughput gain can be achieved if a larger β, and hence increased unfairness, is allowed. Furthermore, in our simulations, flows are always backlogged, a specialized scenario for ease of studying fairness. If otherwise flows are bursty or on/off, the credit of the flow will be reset to zero once the flow becomes unbacklogged. This will alleviate the problem of credit accumulation to a large extent.



(a) Different Maximum Channel Variation



(b) Different Mobility Conditions

Fig. 5. Throughput Gain vs. β

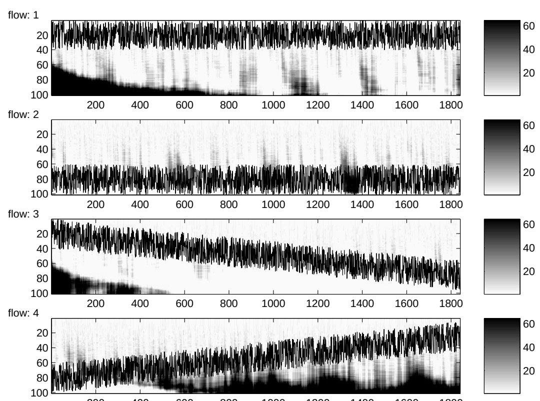


Fig. 6. Dynamic Fairness over β including Channel Conditions

D. Dynamics of Fair Scheduling

In Figure 6, we illustrate the dynamics of WCFQ's fairness. We consider a fixed monitoring window of 4*60=240 time slots and show a subgraph for each flow, depicting fairness in shades of gray, and the experienced channel conditions as a black line. The y-axis of each graph depicts β, which ranges exponentially in 150 steps for 150 simulations $\beta = 0, 10^{[0,0.02,0.04,\cdots,3]}$ from perfect temporal fairness ($\beta = 0$) at the top to $\beta = 1000$ at the bottom. Simulations with larger β are not shown since the fixed window allows a maximum of 60 time slots of unfairness per flow. This window becomes too small for large cost functions and would require separate experiments with larger monitoring windows.

As described in the legend to the right of the graph, the background color depicts the "unfairness" with shades of gray, such that a darker color indicates higher unfairness. Here, unfairness is measured in comparison with the baseline CBFQ schedule. The unfairness value $l = \max((TS_{ref} - TS), 0)$ represents the absolute lack of experienced scheduled time slots TS compared to the CBFQ reference TS_{ref} as monitored within the sliding monitoring window of 240 time slots. Hence, $l = 0$ represents perfect temporal fairness and higher values (up to a maximum of 60 time slots) represent increased unfairness over time.

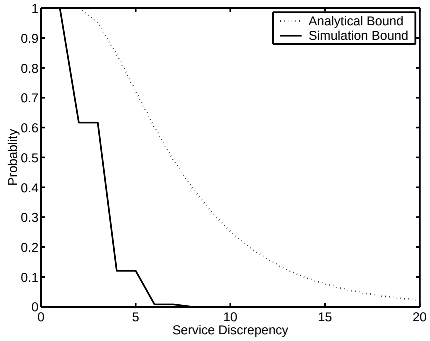
Observe that from simulation 100 downwards on the y-axis, corresponding to $\beta > 100$, there is significant initial unfairness for the two flows 2 and 4 both suffering from poor radio conditions. For larger β values, it is visible that it takes a longer time of 500 to 1000 time slots for these flows to gain sufficient credits to obtain fair temporal channel access. This is reflected in the black slope. After simulation time 600 on the x-axis, user 3 moves towards worse conditions and receives less access, hence the black shades. Note that user 2, which is static and also suffers from poor channel conditions no worse than user 3, has cumulated sufficient credits to receive relatively fair service at this time, even for large cost functions.

E. Comparison with Theoretical Statistical Fairness Bounds

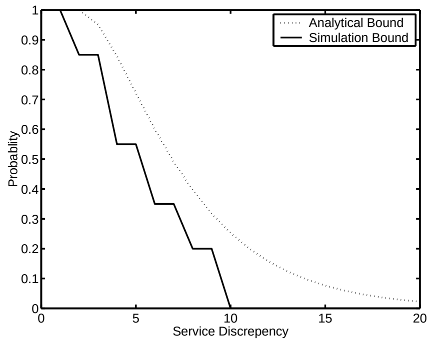
In Section III, we demonstrated that WCFQ provides a statistical fairness bound for any time interval. To compare the theoretical bound with simulation results, we fix the measurement window to 20 packet times and measure the service deviations over non-overlapping windows and compare the distribution with the analytical result. Figure 7 depicts a typical result for a simulation with two flows having equal weights, fixed packet lengths, and uniform and uncorrelated channel conditions generated by using independent uniform distribution in U(0,1) in each time slot.

The dotted line in Figure 7(a) depicts the analytical statistical fairness bound with respect to the service discrepancy for the above cost function with $\beta = 2$. Notice that since we employ the service discrepancy as the x-axis, the probability bound $g(\cdot)$ is $2\exp(-\frac{x/2-1}{2}) - \exp(-2\frac{x/2-1}{2})$. The solid line indicates the probability that the service discrepancy is no less than x. The figure illustrates that the simulation results are well below the analytical bound, indicating that our analytical result indeed bounds the statistical distribution of the service discrepancy, and that this bound is conservative for this particular case.

The tightness of the general bound is closely related to the specific channel model. Intuitively, the more *different* the channels are, the tighter the bound is. The reason is that the difference of the channel provides the chance for WCFQ to exploit the limits of the fairness bound to achieve higher throughput. In the above simulation setup, the channel conditions are uniform i.i.d for each time slot. This means that the unfairness will not be large, even if a scheduler simply selecting the best channel user is employed. To better explore the tightness of analytical result, we change the channel conditions to be more heterogenous and long-time-scale which are generated by triangle waves with random phases. Figure 7(b) compares the simulation and analytical results for this setup. It clearly shows that the analytical bound becomes much tighter in such a scenario. For a particular channel model, algorithms with tighter bound may



(a) Uninform Generation for Each Time Slot



(b) Triangle Wave with Random Phase

Fig. 7. Comparison of Analytical and Simulation Results

be derived by taking into account the special characteristic of the channels. The generality of WCFQ unavoidably costs its tightness.

V. Related Work

A number of previous approaches to wireless fair scheduling are based on the simplified binary channel model [4], [1], [5], [6], [3]. This model assumes either good or bad channel conditions ultimately dictating the scheduling decision to *never* schedule a user in a bad state. Important issues for design of such schedulers include devising compensation strategies to balance throughputs such that lagging flows can catch up after previously experiencing bad channel conditions. However, with adaptive transmission techniques employed in many commercial systems, any user can transmit at any time, yet with different performance and costs attained by different users. Thus, we consider a continuous channel model to exploit this property.

Other more recent work has also considered continuous channels and multi-rate transmission. For example, an adaptive receiver-based scheme is presented in [17] to allow users to obtain throughput gains in IEEE 802.11 systems. However, scheduling and fairness issues are not addressed, so that the scheme does not attain throughput fairness, and has the same temporal fairness properties as IEEE 802.11. In [18] an adaptive algorithm is developed to target different user throughput ratios. With identical throughput targets, the scheme can achieve *asymptotic throughput* fairness [2], whereas WCFQ achieves *short-term temporal* fairness. The objective in [19] is to develop a continuous channel scheduling scheme that maximizes system throughput subject to fairness constraints. While [19], like WCFQ, considers a *temporal* share model of fairness, the scheme targets mainly at *asymptotic* fairness. Although the authors also have developed a scheduling mechanism to enhance short-term fairness (increasing the chances of scheduling lagging flows while limiting the chances for leading flows), it does not quantify the service discrepancy. In contrast, our objective of *provable probabilistic short-term* fairness.

[2]"Asymptotic" is used in the original paper to denote long term fairness. It stems from that convergence is generally required to achieve the fairness.

Another area of research on fair scheduling is rooted in the context of ad hoc networks. For example, in [20] the authors address the fairness issue for ad hoc networks in a distributed environment on the infinite time horizon. However, this kind of research mainly targets at fairness under spatial reuse and randomness of the access protocol opposing to our assumption of centralized control and randomness due to channel condition.

Finally, a set of papers (e.g., [21][22][20]) in the wire line networks and their extensions in the wireless domain (e.g., [23]) have proposed using utility based mechanism to achieve fairness. Though utility is quite similar to the concept of the cost function in this paper, their goal of fairness is also asymptotic and thus clearly different from ours.

Thus, while obtaining fair service in wireless networks is a long standing goal, our work represents the first design of an opportunistic continuous-channel-model scheduler with provable *probabilistic short-term fairness properties*. Such short-term fairness is essential for both delay sensitive applications that are intolerant to short-term service outages, and highly bursty traffic sources in which the duration of a single burst is below the time horizon of the long-term fairness guarantee.

VI. Conclusions

This paper addresses the problem of how to balance fair channel access with opportunistic scheduling of low cost users. We introduced and analyzed WCFQ, the first wireless scheduling algorithm that provides *short term* and *long term* statistical fairness guarantees for a continuous channel model. By considering a general cost function, WCFQ allows system operators to obtain a range of behaviors ranging from perfect temporal fairness to purely opportunistic best-user scheduling. In between, WCFQ provides significant throughput gains while maintaining provable statistical fairness properties.

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] Y. Cao and V. Li, "Scheduling algorithms in broad-band wireless networks," *Proceedings of the IEEE*, vol. 89, no. 1, pp. 76–87, Jan. 2001.

[3] D. Eckhardt and P. Steenkiste, "Effort-limited fair scheduling for wireless networks," in *Proceedings of IEEE INFOCOM '00*, Tel Aviv, Israel, Mar. 2000.

[4] S. Lu, V. Bharghavan, and R. Srikant, "Fair scheduling in wireless packet networks," *IEEE/ACM Transactions on Networking*, vol. 7, no. 4, pp. 473–489, Aug. 1999.

[5] 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.

[6] P. Ramanathan and P. Agrawal, "Adapting fair queuing algorithms to wireless networks," in *Proceedings of ACM MOBICOM '98*, Dallas, TX, Oct. 1998.

[7] 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.

[8] 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.

[9] B. Bensaou, K. Chan, and D. Tsang, "Credit-based fair queueing (CBFQ): A simple and feasible scheduling algorithm for packet networks," in *Proceedings of the IEEE ATM '97 Workshop*, Lisbon, Portugal, 1997.

[10] H. Zhang, "Service disciplines for guaranteed performance service in packet-switching networks," *Proceedings of the IEEE*, vol. 83, no. 10, pp. 1374–96, Oct. 1995.

[11] 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.

[12] 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.

[13] 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.

[14] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Services provided by the physical layer (release 4) ts 25.302 v4.1.0," http://www.3gpp.org, June 2001.

[15] P. Goyal, H. Vin, and H. Chen, "Start-time fair queuing: A scheduling algorithm for integrated services," in *Proceedings of ACM SIGCOMM'96*, Aug. 1996.

[16] A. Parekh and R. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: the single-node case," *IEEE/ACM Transactions on Networking*, vol. 1, no. 3, pp. 344–357, June 1993.

[17] 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.

[18] S. Borst and P. Whiting, "Dynamic rate control algorithms for hdr throughput optimization," in *Proceedings of IEEE INFOCOM '01*, Anchorage, Alaska, Apr. 2001.

[19] X. Liu, E. Chong, and N. Shroff, "Transmission scheduling for efficient wireless network utilization," in *Proceedings of IEEE INFOCOM '01*, Anchorage, Alaska, Apr. 2001.

[20] J. Li, X. Gao, L. Qian, and V. Bharghavan, "Goodput control for heterogeneous data streams," in *Proceedings of the 10th International Workshop on Network and Operating System Support for Digital Audio and Video*, Chapel Hill, NC, June 2000.

[21] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control communications networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–52, Mar. 1998.

[22] S. Shenker, "Fundamental design issues for the future internet," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1176–1188, 1995.

[23] X. Gao, T. Nandagopal, and V. Bharghavan, "Achieving application level fairness through utility-based wireless fair scheduling," in *Proceedings of GLOBECOM'01*.