TECHNICAL RESEARCH REPORT

An Analysis of Delay-Constrained Opportunistic Scheduling for Cellular Wireless Systems

by Roshni Srinivasan, John S. Baras

CSHCN TR 2004-2
(ISR TR 2004-3)

An Analysis of Delay-Constrained Opportunistic Scheduling for Cellular Wireless Systems

Roshni Srinivasan & John S. Baras
roshni@eng.umd.edu, baras@eng.umd.edu
I.S.R., University of Maryland

Abstract

Base station schedulers in 3G and evolving 4G cellular systems use knowledge of the time-varying channel conditions of mobile users to exploit the multiuser diversity inherent in wireless networks. Although such *opportunistic* schedulers significantly improve the system throughput by scheduling users when their channel conditions are most favorable, they could degrade the user experience as a result of unfair resource allocation and increased variability in the scheduled rate and delay. The growing need to provide service differentiation between delay-sensitive multimedia traffic and non real-time data traffic over packet switched air-interfaces underscores the need for these schedulers to incorporate delay constraints.

In this work, we focus primarily on the trade-off between the realization of multiuser diversity gain and the provision of delay guarantees. Our main contribution is an analytical characterization of the distributions of the delay and rate offered by an opportunistic scheduler. The scheduling metric used in the algorithm combines the rate requested by the user and scheduling delay in a general form. Our analysis of a wireless system with a finite number of users in discrete time is strongly supported by system simulations of a time-slotted cellular downlink shared by multiple mobile users with independent, fading channels. We also compute closed form expressions for the scheduler statistics using a continuous approximation. The results in this paper can be used to evaluate system performance and provision resources to support Quality of Service (QoS) guarantees in broadband wireless networks.

I. Introduction

Technologies that provide broadband data services in 3G wireless systems use a combination of circuit-switching and packet-switching for data transmission. Evolving 4G systems are entirely packet-switched networks capable of supporting much higher data rates. In both systems, channel-state dependent scheduling algorithms provide significant gains in the wireless link throughput by exploiting *multiuser diversity*. As the name suggests, multiuser diversity [1] comes from exploiting the fading in the wireless channels of multiple users in the cellular context. Knopp and Humblet [2] showed that in a single cell, where multiple mobile users transmit to a base station, the total uplink information-theoretic capacity is maximized by allowing the mobile with the best channel to utilize the common channel resource. On the downlink, a similar scheduling strategy at the base station has been shown to be maximize throughput [3]. The concept of multiuser diversity has given rise a new class of schedulers in cellular wireless systems that are frequently referred to as *opportunistic* schedulers.

The maximum SNR scheduler best illustrates the idea behind opportunistic schedulers. In a time-slotted system where mobile users constantly report channel quality to the transmitter, this scheduler maximizes system throughput by transmitting to the user with the best channel in every time slot. 3G systems such as the Enhanced Data rates for Global Extension (EDGE) extension in Enhanced General Packet Radio Service (EGPRS) [4] or the High Data Rate (HDR) [5], [6] extension of CDMA2000 also give priority to users with better channels by utilizing the forward-link channel state information reported by mobile users. It is easy to see that the gains in system throughput come at the cost of unfair resource allocation and variability in the scheduled rate and delay. In order to support a traffic with a wide range of Quality of Service (QoS) requirements, these schedulers must incorporate delay constraints.

Scheduling algorithms for wireless networks that are optimized to support delay guarantees [7], [8], [9], [10], [11], [12] have been well studied in the literature. Feasibility and complexity limit the practical implementation of these schedulers in providing explicit QoS guarantees. In addition to the design of an efficient scheduling algorithm, the evaluation of its performance in achieving the desired throughput or delay is of great importance to a network operator. To the best of our knowledge, the results in this paper are first to completely characterize the scheduled rate and delay experienced by mobile users in terms of a configurable scheduler metric. The metric used has a more general form, combining an estimate of the user's channel with the scheduling delay experienced by a user. Furthermore, the analysis highlights the inherent trade-off between system throughput and the delay experienced by mobile users with opportunistic scheduling. The results in this paper address the important issue of providing QoS suport in cellular wireless systems. Quantifying the performance at the MAC layer also benefits higher layer protocols such as TCP which are impacted by rate and delay fluctuations [13], [14]. This cross-layer networking approach can further enhance system performance.

The outline of the paper is as follows. In Section II, we illustrate the trade-off between multiuser diversity gain and scheduling delay in opportunistic scheduling. Section III discusses related work in this area of research. We develop an analytical framework for the calculation of scheduled rate and delay distributions in Section IV. Section V contains system details and implementation issues. Simulation results that validate our analysis are presented in Section VI. Finally, we discuss the main ideas in this work in our conclusions in Section VII.

II. Motivation

The gains from opportunistic scheduling come at the expense of scheduling delay and delay jitter. In order to illustrate this trade-off, consider the following scheduler metric, $m(t)$, that combines multiuser diversity gain with delay constraints:

$$m(t) = R(t) + \alpha\frac{v(t)}{N} = R(t) + \alpha V(t), \tag{1}$$

where $R(t)$ is the rate requested by the mobile at the beginning of time slot $[t, t+1)$, $v(t)$ is the delay since a waiting packet in the mobile user's queue was previously served and α is a configurable control weight. The scheduling delay $v(t)$ at the beginning of time slot $[t, t+1)$, normalized by the number of users is represented by $V(t)$. In the case of opportunistic schedulers, waiting packets could be delayed because the scheduler is serving other users either because their channel conditions are better or for reasons of fairness. We therefore refer to the normalized delay, $V(t)$ as scheduler *vacation time* in the rest of this paper. As the number of users increases, $v(t)$ increases proportionally. Using the normalized version of $v(t)$ ensures that the number of users does not affect the balance between multiuser diversity gain and delay implicit in the metric.

In Figure 1, we plot the coefficient of variation (ratio of standard deviation to the mean) of the scheduled rate and the coefficient of variation (CV) of the delay between scheduling slots as a function of α. The scenario we consider is the downlink of a cellular wireless system similar to the 1xEV-DO [5], [6] data system. In such a system, a base station serves N mobile users in a cell using a time-slotted downlink combined with an asynchronous circuit-switched uplink. In this example, we consider 16 users, each with a nominal SNR of 2.5dB. We assume a Rayleigh SNR distribution for the flat fading channel experienced by every user. The maximum sustainable rate, which is a function of the channel quality is constantly reported back to the base station by every mobile via a dedicated, circuit-switched channel on the reverse link. These *requested rates* are independent of each other and across time slots, but identically distributed. In every time slot, the scheduler at the base station computes a metric for each user as given

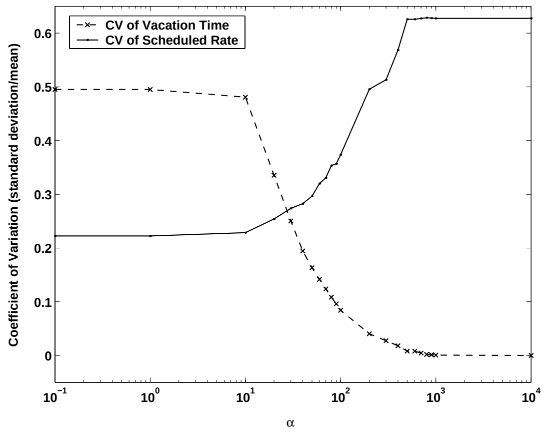

Fig. 1. CV of Scheduled Rate and CV of Vacation Time vs alpha for 16 Users at Nominal SNR of 2.5dB

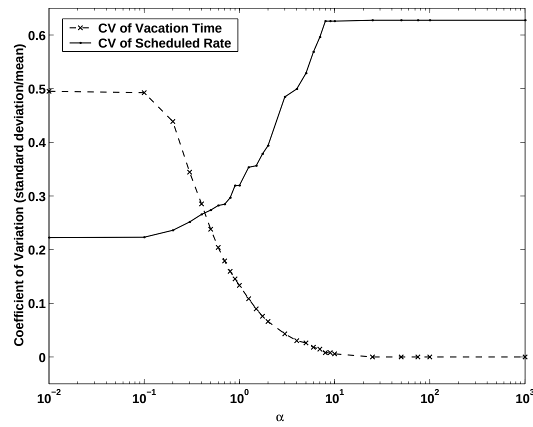

Fig. 2. CV of Scheduled Rate and CV of Vacation Time vs alpha for 16 Users at Nominal SNR of 2.5dB with Scaled Rate

by equation 1. The base station then transmits to the user with the highest metric.

We see from Figure 1 that for small values of α, the scheduler described above behaves like the Maximum SNR scheduler. The mean scheduled rate is large since the scheduler always serves the user with the best channel and consequently, the standard deviation of the scheduled rate is also very low. The CV of the scheduled rate is therefore the lowest for the Maximum SNR scheduler. On the other hand, the scheduling delays are given no priority, causing large variations in the vacation time experienced by users and therefore a high coeffecent of variation. As α is increased, the contribution of the delay towards the scheduling metric increases. For large α , the scheduler is channel agnostic. Users are served cyclically, once in every N slots as in Longest Wait First (LWF) scheduling. The LWF scheduler, naturally has the lowest CV of vacation time. The scheduled rate exhibits the opposite trend. We measure the vacation time in slots and the scheduled rate in Link Layer segments per slot. Table I lists the transmitted rate in link-layer segments as a function of the SNR. For 8 byte segments and a slot duration of 1.667ms, the maximum rate corresponds to 64 segments per slot. Although the dynamic ranges of the rate and normalized vacation time in the metric proposed in equation 1 are not of the same scale, normalizing the scheduled rate by the maximum rate of 64 segments resolves the issue of dimensionality. Figure 2 demonstrates that this normalization simply corresponds to a scaling in the range of α over which the scheduled rate and vacation time statistics vary.

The trade-off between system throughput and scheduling delay can also be seen in the metric used to select a user in the PF scheduler [15]

$$m(t) = \frac{R(t)}{T(t)}, \tag{2}$$

where $T(t)$ is an estimate of the user's average MAC layer throughput in some window of time prior to the current instant. If i^* denotes the user with the highest metric, then $T(t)$ can be obtained by exponentially averaging the ith user's throughput over a scheduling time scale t_c i.e.,

$$\begin{aligned} T_i(t+1) = &\quad (1-\tfrac{1}{t_c})T_i(t) + (\tfrac{1}{t_c})R_i(t), \qquad i = i^* \\ = &\quad (1-\tfrac{1}{t_c})T_i(t) \qquad\qquad i \neq i^* \end{aligned} \tag{3}$$

The parameter t_c is tied to the latency time scale and the QoS requirement of the application. If the latency time scale is large then the scheduler has more flexibility in scheduling the user, possibly waiting much longer until the user's channel is at a very high peak. While the scheduled rate is bound to be higher, the variations in scheduling delay are larger. When $t_c \to \infty$, the long-term average throughput of each user is shown to exist and the algorithm is shown to maximize $\sum_{i=1}^{N} logT_i$ among the class of all schedulers. The PF scheduler therefore provides an implicit mechanism to increase aggregate cell throughput at the MAC layer at the expense of increased variability in scheduling delay jitter.

III. Related Work

Schedulers for wireless networks, [16], [17], [18] were first designed along the lines of their fair queueing analogs in wireline systems. Channel-state-dependent schedulers that enhance throughput have been proposed for wireless LANs [19], [20]. Knopp and Humblet [2] were among the first to recognize the gains realizable in a cellular wireless system with multiple users by scheduling users when their channel conditions are favorable. The Proportional Fair scheduler used in 1xEV-DO [15] is the first application of an opportunistic scheduler in a commercial wireless system. In recent years, improving wireless technologies have increased the demand for wireless data services [21]. Real-time applications such as voice, video conferencing and games share network resources with non-real-time traffic such as file transfers and messaging. QoS support for wireless data is therefore a natural consequence of the integration of packet-switched wireless networks with the internet, placing new demands on scheduling algorithms.

The low delay scheduling algorithm proposed by Bettesh and Shamai [7] was one of the first to address the problem of arbitrarily long delays that could result from opportunistic scheduling. The theoretically complex evaluation of the performance of this algorithm limited its application. The Modified - Largest Weighted Delay First (M-LWDF) [9] rule attempts to optimally provide QoS guarantees in terms of predefined guarantees for the probability of loss and minimum long-term throughput for each user. The exponential rule [10], [11] is optimized to share a time-varying channel among multiple real-time users with deadlines.

A framework that provides QoS guarantees has two important elements. The first is an admission control policy that determines the extent to which a user can utilize system resources. The second is a mechanism to evaluate the QoS obtained by an appropriate scheduling algorithm in terms of its control parameters. In effect, this mechanism determines whether the QoS requirements of an admitted user can be supported. In this paper, we address the latter issue by analytically computing the distributions for the scheduled rate and resultant delay as a function of the control parameters in the scheduler metric. The problem is complicated by the time-varying wireless channel capacity as a result of fading. The authors in [22] introduce the notion of effective capacity to address this issue. Using effective capacity for for admission control, they design a scheduling algorithm that combines Round Robin Scheduling with Maximum SNR scheduling to provide QoS guarantees . We take a new approach and model the system formed by the scheduler at the base station and the mobile users as a dynamical system. In the discrete case, the analysis models a time-slotted system with a finite number of users. We also approximate the behaviour of the system in continuous time and compute closed-form expressions for the scheduler statistics.

IV. Analysis

We now analyze the system comprising of the delay-constrained opportunistic scheduler proposed in Section II and the time-varying channel conditions of multiple mobile users as a dynamical system. The analysis outlined here characterizes the distribution of the delay experienced between successive scheduling instants and the distribution of the scheduled rate. The approach used is to consider the scheduler as a dynamical system and to then examine this system in its steady-state.

A. The Discrete Case : A Time-slotted System with a Finite Number of Users

The analysis in this section rests on the following assumptions. The base station shares the downlink among a finite group of N users with identical channel statistics. Let $\mathcal{N} = \{0, \ldots, N-1\}$ denote the set of mobile users served by the base station. Each user $i \in \mathcal{N}$ indicates the maximum sustainable rate to the base station on a dedicated channel on the uplink. Let $R_i(t)$ denote the rate requested by user i at the beginning of time slot $[t, t+1)$. In order to make the analysis more tractable, the channel rates for each user are assumed to be independent from one time slot to the other. Naturally, the requested rates of the users are independent of each other, but identically distributed. Let $\mathcal{R} = \{r_0, r_1, \ldots, r_{max}\}$ denote the finite set of rates requested by the users. This set is assumed to have a probability distribution $f_{R_i}(r) = f_R(r) = P(R = r), r \in \mathcal{R}, \forall i \in \mathcal{N}$. The delay experienced by each user $i \in \mathcal{N}$ since it was previously scheduled is $v_i(t)$, with $V_i(t) = v_i(t)/N$ representing the normalized vacation time at the beginning of time slot $[t, t+1)$.

In every time slot, the base station transmits to the user with the highest metric computed from equation 1, applying a tie-breaking rule if necessary. The analysis of the scheduler in the state space formed by the users, their delays and time-varying channel conditions is very complex. However, the problem becomes tractable and easily amenable to analysis when the state space is defined based on a permutation of the user space in which the users are rank-ordered in every slot according to the delay they have experienced since they were last scheduled. Let $\mathbf{U}(t)$ denote the rank-ordering of users at the beginning of time slot $[t, t+1)$.

$$\mathbf{U}(t) = \{u_0(t), u_1(t), ..., u_{N-1}(t)\}, \tag{4}$$

where $u_i(t) \in [0, 1, ..., N-1]$. In this space, $u_i(t)$ denotes the original index of the user who is ranked in the i^{th} position at at the beginning of time slot $[t, t+1)$. By definition, this permutation has the property,

$$V_{u_0}(t) \leq V_{u_1}(t) \leq ... \leq V_{u_{N-1}}(t) \tag{5}$$

where $V_{u_i}(t)$ is the vacation time seen by the user who is ranked in position i at the beginning of time slot $[t, t+1)$. Naturally, since $u_0(t)$ is the index of the user scheduled in the previous time slot, $V_{u_0}(t) = 1/N$. At the beginning of time slot $[t, t+1)$, the scheduler selects a user whose rank, $S^*(t)$ is given by

$$S^*(t) = \arg\max_i m_{u_i}(t), \quad i \in \mathcal{N} \tag{6}$$

In the event that more than one user has the highest metric, $S^*(t)$ is picked with uniform probability from among the users with the highest metric in order to break the tie. The selection of a user in one slot causes the rank-ordering of the users to change at the beginning of the next slot. At the beginning of time slot $[t+1, t+2)$, the user, S^* that was selected in the previous time slot moves to position 0 in the rank-ordered space. Since users are arranged in ascending order of their vacation times, all users with rank greater than that of S^* do not change their order in any way, while all users below the rank of S^* increment their rank by one. Specifically, $\mathbf{U}$ evolves over time as:

$$u_i(t+1) = \begin{cases} u_{S^*}(t), & i = 0 \\ u_i(t), & i = S^*(t)+1, S^*(t)+2, \ldots, N-1 \\ u_{i-1}(t), & i = 1, 2, \ldots, S^*(t)-1 \end{cases} \tag{7}$$

Correspondingly, the vacation time seen by every user who was not scheduled increases, while the vacation time seen by the scheduled user is reset to the minimum possible value

$$V_{u_i}(t+1) = \begin{cases} \frac{1}{N}, & i = 0 \\ V_{u_{i-1}}(t) + \frac{1}{N}, & 0 < i < S^*(t) \\ V_{u_i}(t) + \frac{1}{N}, & i > S^*(t) \end{cases} \tag{8}$$

This framework describes the evolution of a dynamical system consisting of the rank-ordered user space, the corresponding channel conditions and scheduling delays.

We define a selection density function, $\pi_{u_i}(t)$ which represents the probability of scheduling the ith rank-ordered user, u_i at the beginning of time slot $[t, t+1)$.

$$\pi_{u_i}(t) = Pr(S^*(t) = u_i), \quad i \in \mathcal{N} \tag{9}$$

with the property, $\sum_{i=0}^{N} \pi_{u_i}(t) = 1$. The notion of a selection density function is better understood through the following examples. First, consider a Longest Wait First (LWF) scheduler, which always serves the queue with the largest vacation time. This corresponds to a choice of $\alpha \to \infty$ in the composite metric in Equation 1. In this case, the selection density function may be written as

$$\pi_{u_i}(t) = \begin{cases} 0, & i = 0, 1, \ldots, N-2 \\ 1, & i = N-1 \end{cases} \tag{10}$$

Contrast this with a Maximum SNR scheduler, which by definition is agnostic to the delay experienced by any user. This scheduler results from a choice of $\alpha = 0$ in Equation 1. If the N users have identical channel statistics, they are equally likely to be served, which results in a uniform selection function

$$\pi_{u_i}(t) = \frac{1}{N}, \quad i \in \mathcal{N} \tag{11}$$

The vacation function $V_u(t)$ and the selection density function $\pi_u(t)$ may be composed with each other to derive statistical measures of the vacation time. For instance, the k^{th} moment of the vacation time seen by the user population at the beginning of time slot $[t, t+1)$ is given by

$$E[V_u^k(t)] = \sum_{i=0}^{N-1} V_{u_i}^k(t) \pi_{u_i}(t) \tag{12}$$

Hence, knowledge of the functions, $V_u(t)$ and $\pi_u(t)$ for any choice of α in the metric completely determines the distributions of the vacation time between scheduling instants as well as the scheduled rate, thereby fully characterizing the scheduler. An iterative computation of $V_u(t)$ and $\pi_u(t)$ is used to obtain the fixed-point, time-invariant solutions of the dynamical system, i.e., V_u and π_u. In the following sections, we compute the functions, V_u and π_u through analysis.

1) Computation of the Selection Density function π_u: At the beginning of time slot $[t, t+1)$, the user with rank-ordered index i has a vacation time of $V_{u_i}(t)$. If $R_{u_i}(t)$ is the user's requested rate, the metric $m_{u_i}(t)$ for each user $i \in \mathcal{N}$ is given by

$$m_{u_i}(t) = R_{u_i}(t) + \alpha V_{u_i}(t)$$

The probability of selecting the i^{th} user is then given by

$$\begin{aligned} \pi_{u_i}(t) &= P(m_{u_i}(t) > m_{u_j}(t) \quad \forall j \neq i) + P(u_i \text{ is selected in a tie}) \\ &= P(R_{u_i}(t) + \alpha V_{u_i}(t) > R_{u_j}(t) + \alpha V_{u_j}(t) \quad \forall j \neq i) + P(u_i \text{ is selected in a tie}) \end{aligned}$$

We employ a simple tie-breaking rule in the event that more than one user has the highest metric. In such a case, a single user is picked with uniform probability from among the set of users with the highest metric. The computation of the probability of selecting the i^{th} user in the event of a tie is outlined in the Appendix. Since the channel rates are i.i.d. random variables with distribution $f_R(r)$,

$$P(m_{u_i}(t) > m_{u_j}(t)) = \sum_{r=r_0}^{r_{max}} \prod_{j \neq i} F_R(r + \alpha(V_{u_i}(t) - V_{u_j}(t))) f_R(r)$$

where r_{max} is the maximum rate that can be supported by the mobile user. In the steady state, the selection density function π_{u_i} for each user $i \in \mathcal{N}$ is given by

$$\pi_{u_i} = \sum_{r=r_0}^{r_{max}} \prod_{j \neq i} F_R(r + \alpha(V_{u_i} - V_{u_j})) f_R(r) + P(u_i \text{ is selected in a tie}), \tag{13}$$

where V_{u_i} represents the vacation function for user $i \in \mathcal{N}$ at the fixed-point of the system.

2) Computation of the Vacation Function, V_u: The vacation function, V_u characterizes the normalized delay experienced by the users in the system. In the analysis that follows, we assume the existence of a selection density function, $\pi_{u_j}(t)$ which represents the probability of scheduling the jth rank-ordered user, u_j at the beginning of time slot $[t, t+1)$.

Recall from equations 7 and 8 that in the set of rank-ordered users, $\mathbf{U}$, the selected user S^* at the beginning of time slot $[t, t+1)$ moves to position 0 at the beginning of time slot $[t+1, t+2)$. All users with rank greater than S^* experience an increase in delay, but do not change their rank in any way. Therefore, if $S^*(t) = \arg\max_i m_{u_i}(t)$, the vacation time for these users evolves as

$$V_{u_i}(t+1) = V_{u_i}(t) + \frac{1}{N}, \quad i > S^*(t)$$

As a result of the scheduled user occupying the very first position in the next slot, all users with rank less than S^* experience an increase in delay and also an increase in rank by one.

$$V_{u_{i+1}}(t+1) = V_{u_i}(t) + \frac{1}{N}, \quad i < S^*(t)$$

Therefore, the vacation function at position i in the rank-ordered space is subject to two transforming forces. The first causes its value to increase by $1/N$ whenever a user with a rank less than i is scheduled. This event occurs with probability $\sum_{j<i} \pi_{u_j}(t)$. The second transformation causes its value to decrease whenever the rank of the user scheduled is i or higher. In this event, the value of the vacation-time at position i is replaced by that at position $(i-1)$, augmented by $1/N$. The probability of this event is $\sum_{j \geq i} \pi_{u_j}(t)$.

In an equilibrium state, the vacation function is invariant to these two transforming forces with the potential increase balancing the potential decrease. Dropping the dependence on time, we then have,

$$\frac{1}{N}\left(\sum_{j<i} \pi_{u_j}\right) = \left(V_{u_i} - \left(V_{u_{i-1}} + \frac{1}{N}\right)\right)\left(\sum_{j \geq i} \pi_{u_j}\right)$$

The vacation function V_u at equilibrium may be computed recursively as

$$V_{u_i} = V_{u_{i-1}} + \frac{1}{N\left(1 - \sum_{j<i} \pi_{u_j}\right)}, \quad i = 1, \ldots N-1 \tag{14}$$

with the initial condition, $V_{u_0} = \frac{1}{N}$.

3) Worst-case Normalized Delay: The upper bound on the normalized delay seen by any user can be computed from equation 14 as

$$V_{max} = V_{u_{N-1}} = \frac{1}{N} + \sum_{j=1}^{N-1} \frac{1}{N\left(1 - \sum_{j<i} \pi_{u_j}\right)} \tag{15}$$

Observe that the selection density function completely determines this upper bound. The LWF scheduler, with $\alpha \to \infty$ has the lowest worst-case normalized delay since the selection density function is 1 for the highest rank-ordered user and 0 for all other users. No term in the summation in equation 15 is larger than $1/N$. On the other hand, in opportunistic schedulers for which α is very small, users with lower delays but better channel conditions have a higher proability of being scheduled. The selection density functions for such schedulers have non-zero values for rank-ordered users with lower delays in addition to those with higher delays. The term in the summation in equation 15 is larger than $1/N$ for many users, thereby causing the maximum normalized delay to increase as α is decreased. The Maximum SNR scheduler naturally has the highest normalized vacation time.

4) Distributions for Scheduled Rate and Vacation Time: Given the fixed-point of the dynamical system, the distribution of vacation time at the scheduling instants can be obtained easily from the functions, V_u and π_u . Let $\mathcal{V}_{S^*}$ denote the random variable representing the vacation time seen by the *scheduled* user. For some non-negative number γ,

$$\begin{aligned} P[\mathcal{V}_{S^*} \leq \gamma] &= \sum_{j=0}^{i(\gamma)} \pi_{u_j}, \quad \text{where} \\ i(\gamma) &= \arg\max_k (V_{u_k} \leq \gamma) \end{aligned} \tag{16}$$

The pdf of the scheduled rate (which is naturally different from that of the requested rate) may be derived as a function of α as

$$\begin{aligned} f_{R_{S^*}}(r) &= \sum_{i=0}^{N-1} Pr(R_{u_i} = r, i\text{th rank-ordered user is selected}) \\ &= \sum_{i=0}^{N-1} f_R(r) \prod_{j \neq i} F_R(r + \alpha(V_{u_i} - V_{u_j})) + Pr(\text{User } i \text{ is selected in a tie}) \end{aligned} \tag{17}$$

B. Continuous Approximation for the Vacation Function

In this section, we derive an approximation for the vacation function in continuous time as the number of users, $N \to \infty$. The normalized, rank-ordered user space is now defined over the continuum $[0, 1]$ and the duration of a slot is infinitesimally small. Let the function $V(u, t)$ represent the vacation time experienced by a user with rank-ordered index, $u \in [0, 1]$ at time t. At any given time t, a fraction of users, u will see a vacation time that is no larger than $V(u, t)$. Naturally, $V(0, t) = 0$ since this is the vacation time of the user who has just been served. This function is the continuous analog of V_{u_j}, which in turn may be derived by sampling $V(u, t)$ uniformly in $u \in [0, 1]$ at N points.

We also define a continuous analog of the selection density function, $\pi(u, t)$, which represents the time-varying probability of being scheduled as a function of $u \in [0, 1]$. At time t, the probability of a user mass around the point u being scheduled in a time interval dt is given by $\pi(u, t)dt$, with

$$\int_u \pi(u, t) du = 1 \tag{18}$$

This analog of the selection density function in the discrete case has a familiar interpretation. First, consider a strict LWF scheduler, which always serves the queue with the largest vacation time. This corresponds to a choice of $\alpha \to \infty$ in the composite metric in Equation 1. In this case, it is straightforward to observe that the selection function is an impulse at $u = 1$, with zero weight everywhere else.

$$\pi(u, t) = \delta(u - 1), \quad \forall u \in [0, 1], \quad \forall t \tag{19}$$

Contrast this with a Maximum SNR scheduler, which by definition is agnostic to the delay experienced by any user. This scheduler results from a choice of $\alpha = 0$ in equation 1. If the users have identical channel statistics, they are equally likely to be served, which results in a uniform selection density function

$$\pi(u, t) = 1, \quad \forall u \in [0, 1], \quad \forall t \tag{20}$$

Now, assuming the existence of a selection density function, $\pi(u, t)$, consider the time-evolution of the vacation function, $V(u, t)$. The process of scheduling users constantly subjects the function $V(u, t)$ to transformations. To understand this process, let us focus attention on the partial derivative $V_u'(u, t) = \frac{\partial V(u,t)}{\partial u}$. This derivative is subject to two transforming forces.

Recall from equations 7 and 8 that in the set of rank-ordered users, $\mathbf{U}$, the selected user at time t, $S^*(t)$, moves to position 0 at time $t + 1$. All users with ranks higher than $S^*(t)$ experience an increase in delay, but they do not change their order in any way. As a result of the scheduled user occupying the very first position, all users with ranks less than $S^*(t)$ experience an increase in delay as well as rank. Therefore, the local neighborhood of any point which has a non-zero probability of being scheduled experiences an increase in slope. This corresponds to the fact that the local user mass in any interval du around the point u is reduced due to scheduling. The probability mass of scheduling a user mass at u at time t in an interval dt is $\pi(u,t)dt$. The new user mass in the neighborhood of u is therefore $du[1-\pi(u,t)dt]$. The local slope gets transformed as

$$\begin{aligned} V_u'(u,t+dt) &= \frac{V_u'(u,t)}{(1-\pi(u,t)dt)} \\ &= V_u'(u,t)(1+\pi(u,t)dt), \quad \text{neglecting higher order terms} \end{aligned}$$

The potential increase in the local slope due to this transformation is

$$V_u'(u,t+dt) - V_u'(u,t) = V_u'(u,t)\pi(u,t)dt \tag{21}$$

The curve $V(u,t)$ constantly experiences a transforming force pushing it to the right i.e., increasing u, as a result of the total user mass in the interval $[u,1]$ which gets scheduled. The amount by which the curve is shifted in a time interval dt is given by

$$\begin{aligned} \epsilon &= [\int_{x=u}^{1} \pi(x)dxdt] \\ &= I(u)dt \end{aligned} \tag{22}$$

Hence, the new slope at position u and time $t+dt$ is related to the slope at position $u-I(u)dt$ and time t as

$$V_u'(u,t+dt) = V_u'((u-I(u)dt),t)$$

Expanding $V_u'((u-I(u)dt),t)$ around the point u in a Taylor series and neglecting the higher order terms, we get

$$V_u'((u-I(u)dt),t) = V'(u,t) - V_u''(u,t)I(u)dt$$

where $V_u''(u,t)$ is $\frac{\partial^2 V(u,t)}{\partial u^2}$. The potential decrease in the local slope due to this transformation is given by

$$V_u'(u,t) - V_u'((u-I(u)dt),t) = V_u''(u,t)I(u)dt \tag{23}$$

In an equilibrium state, when the curve is invariant with respect to time, these two transforming forces are equalized. Hence, from Equations 21 and 23, we get

$$V_u''(u,t)I(u)dt = V_u'(u,t)\pi(u,t)dt$$

which, by dropping the dependence on time, may be expressed as

$$\frac{V_u''(u)}{V_u'(u)} = \frac{\pi(u)}{I(u)} = \frac{\pi(u)}{\int_{x=u}^{1} \pi(x)dx}$$

Integrating both sides with respect to u,

$$\begin{aligned} log(V_u'(u)) &= \int_u \frac{\pi(u)du}{\int_{x=u}^{1} \pi(x)dx} \\ V(u) &= \int_u exp(\int_u \frac{\pi(u)du}{\int_{x=u}^{1} \pi(x)dx})du \end{aligned} \tag{24}$$

The vacation function in the steady state, $V(u)$ can therefore be computed from the selection density function in the steady state, $\pi(u)$.

V. System Model and Implementation Issues

In this section we describe the system model and the wireless channel model. We also discuss various aspects of the implementation of the scheduler that uses the metric introduced in Equation 1.

A. System Model

In our simulation experiments, we use a system architecture that is similar to the 3G CDMA wireless data systems such as 1xEV-DO and 1xEV-DV. We combine a time-slotted downlink with an asynchronous cicrcuit-switched uplink. Since most data applications are fundamentally asymmetric and very little data flows on the dedicated uplink, we focus our attention on downlink scheduling alone. We now outline some important features in the system model.

TABLE I
Transmission rate per slot as a function of SNR

SNR (in dB)	Rate (Kb/s)
-12.5	38.4
-9.5	76.8
-6.5	153.6
-5.7	204.8
-4	307.2
-1.0	614.4
1.3	921.6
3.0	1228.8
7.2	1843.2
9.5	2457.6

Packet streams for individual users are assigned separate queues by the base station. Fixed length packets of 512 bytes are segmented into link-layer (LL) segments of 8 bytes for transmission over the air link. At the beginning of each time slot, the scheduler at the base station computes the metric as in equation 1 and selects the data user with the highest metric. The number of segments transmitted in a slot depends on the current SNR of the selected data user; this correspondence is enumerated in Table I. The slot duration of 1.667ms and peak rate of 2.45Mbps achievable in this model are similar to the 1xEV-DO system. When all the LL segments corresponding to the packet at the head of the queue for a particular user have been transmitted over the airlink, the packet is dequeued. Transmission errors can be simulated by probabilistically delaying packet transmission. Since we assume that the channel state is known to a high degree of accuracy, we assume a negligible loss probability. Every user is always assumed to have data in the queue. This ensures that the scheduling metric is the sole criterion for selecting a user.

B. Wireless Channel Model

Under a flat fading assumption, the channel response for every mobile user is assumed to be constant over the duration of the slot. If $x_i(t)$ and $y_i(t)$ denote the vectors of transmitted and received symbols for user i at the beginning of time slot $[t, t+1)$, then

$$y_i(t) = h_i(t)x_i(t) + z_i(t), \quad i = 1, 2, \ldots, N \tag{25}$$

where $h_i(t)$ is the time-varying channel response from the BS to the mobile and $z_i(t)$ is a noise vector with i.i.d., zero mean Gaussian noise components with variance σ_i^2. Assuming unit-energy signals, the nominal SNR for user i is $C_{NOM,i} = \frac{1}{\sigma_i^2}$ with the instantaneous SNR for this user, $C_i(t)$ given by $C_i(t) = \frac{h_i(t)}{\sigma_i^2}$. The probability distribution for the rates requested by the mobile users is generated by using the Jakes [23]

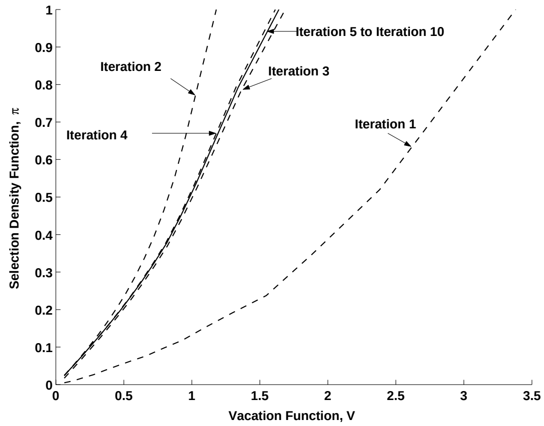

Fig. 3. Iterative Computation of V_u and π_u

model to simulate time-varying channels. The Jakes model uses a sum of K complex exponentials to approximate a single-path Rayleigh fading channel. The complex channel gain at time t is given by

$$h_i(t) = \sum_{j=0}^{K-1} h_{i,j} exp(j2\pi f_d^i t cos(2\pi\phi_j)) \tag{26}$$

where $h_{i,j}, j = 0, .., K-1$ are complex, unit variance gaussian random variables with zero mean representing the magnitudes of the subpaths. Each subpath has a phase delay, ϕ_j, which is uniformly distributed in $[0, 2\pi]$. The doppler frequency of the user is given by f_d^i. The Jakes model produces a sequence of attenuation coefficients that is very close to a Rayleigh fading process, and in particular has the same correlation properties.

We study the performance of 16 users ($N = 16$), all with $C_{NOM,i} = 2.5dB$ and a doppler frequency of 10Hz. A scenario with identical channel statistics for all users was selected to enable comparison between analysis and simulation. For the case of i.i.d. channel fades, the distribution of the channel rates is chosen to be identical to the marginal distribution obtained with correlated channel fades at the same nominal SNR.

VI. Simulation Results

The distributions of the scheduled rates and vaction times can be obtained analytically when the vacation function V_u and the selection density function π_u are known. In this section, we first describe the iterative method used to compute these functions. We then highlight some properties of the vacation function that can be verified through simulations. Finally, we present a comparison of analytical and simulation results for the distribution of the scheduled rates and vacation times parametrized by α.

A. Numerical Computation of V_u and π_u

The selection density function π_u and the vacation function V_u can be computed analytically, assuming knowledge of each other, as outlined in IV-A.1 andSections IV-A.2. In this section, an iterative approach is employed to compute these functions for a scheduler parameterized by any choice of α. Since neither of the functions is known at the outset, we use the following approach. We start with the Maximum SNR scheduler. We see from equation 11 that the selection density function is uniformly distributed among the n users. Let $\pi_u^{(k)}$ and $V_u^{(k)}$ represent the selection density function and the vacation functions estimated in the k^{th} iteration respectively. The maximum SNR scheduling assumption in the first iteration implies that $\pi_u^{(0)} = \frac{1}{N}$. $V_u^{(1)}$ can be therefore be computed using the expression derived in Equation 14.

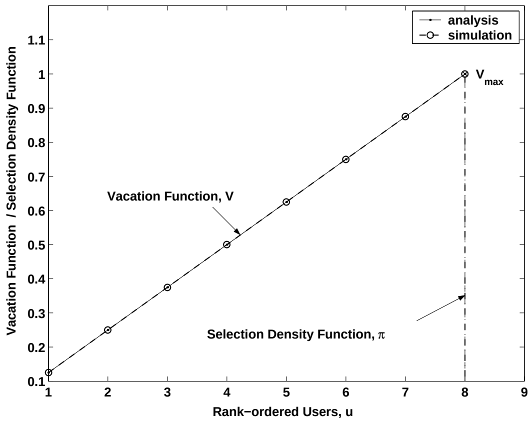

Fig. 4. LWF Scheduler : V and π for $\alpha = 1000$

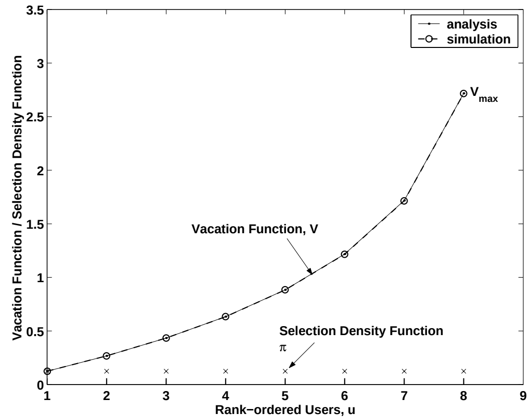

Fig. 5. Maximum SNR Scheduler : V and π for $\alpha = 0$

Correspondingly, $\pi_u^{(1)}$ is computed from $V_u^{(1)}$ using the approach outlined in Section IV-A.1. In subsequent iterations, $V_u^{(k)}$ is computed from $\pi_u^{(k-1)}$, which in turn facilitates computation of $\pi_u^{(k)}$. The convergence of this process has been observed empirically as in Figure 3. We are currently working on a formal proof.

B. The Worst-Case Normalized Delay

The local slope of the vacation function at any position i in the rank ordered user space is given by

$$\Delta V_{u_i} = V_{u_i} - V_{u_{i-1}} \tag{27}$$

Observe from equation 14 that ΔV_{u_i} is given by $1/\left(N\left(1 - \sum_{j<i} \pi_{u_j}\right)\right)$, and is therefore a montonically increasing function of the rank index. The slope at a rank-ordered index i, ΔV_{u_i} is constant $(1/N)$ as long as the probability of scheduling users with lower vacation times is zero i.e., $\pi_{u_j} = 0, \forall j < i$. With increasing i, as the cumulative probability of scheduling, $\pi_{u_j}, \forall j < i$, increases, ΔV_{u_i} increases as well. For every non-zero probability mass in the selection density function, there is a corresponding increase in the local slope of the vacation function. This increase in slope is captured by the summation in the second term of equation 15. The interdependence of the vacation function and the selection density function in the proposed scheduling metric gives rise to implicit upper and lower bounds for the worst-case scheduling delay.

At one extreme, we have the channel agnostic LWF scheduler in which the delay constraint dominates. Fig. 4 shows the vacation function and the selection density function for a choice of $\alpha = 1000$ and 8 users. This approximates the behaviour for the LWF scheduler with $\alpha \rightarrow \infty$. As expected, the vacation function, V_u is linear, with slope $= 0.125$ and the maximum normalized delay, $V_{max} = 1$ both analytically and from the simulations. The selection density function has all the probability mass concentrated at 1, since the scheduler always schedules the user with the highest delay.

At the other extreme, we have the Maximum SNR scheduler in which the channel conditions dominate the metric with no constraint on the delay. As in Fig. 5, the selection density function is shown to be uniform (0.125). The slope of the vacation function increases at every point in the user space. From equation 15, the worst case normalized delay is $V_{max} = 2.7179$. The simulation results for the mean normalized delay experienced by the rank-ordered users closely match the analysis with $V_{max} = 2.7157$.

C. Distributions for Scheduled Rates and Vacation Time

For the purpose of clarity, we plot the CDFs of the vacation time and scheduled rates in Figures 6 & 7 resp. for three values of α. The graphs illustrate the close correspondence of the simulation results

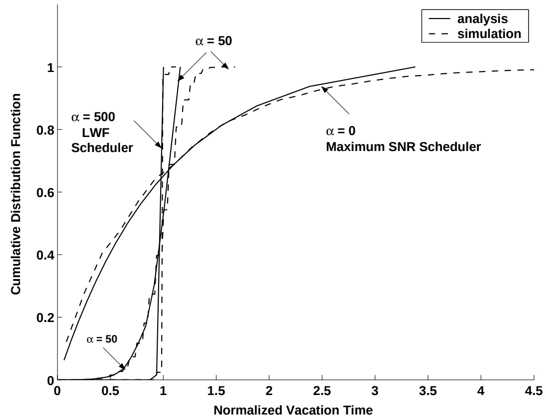

Fig. 6. Vacation Time distribution

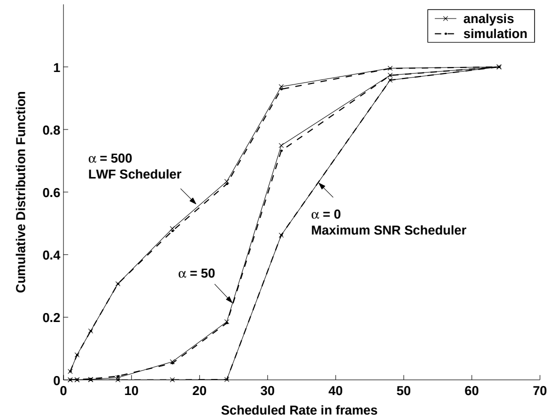

Fig. 7. Scheduled Rate distribution

with the analysis in Section IV. In Figure 6, we see both from simulation and analysis that the LWF scheduler ($\alpha = 500$), concentrates the mass of the CDF at the normalized delay of 1, scheduling users in a Round Robin manner. Since the scheduler is channel agnostic, the probability of a user being selected is zero for all but the user with the highest delay. The Maximum SNR scheduler ($\alpha = 0$), on the other hand, ignores delay, always favoring users with higher SNR. The highest normalized vacation time is almost 5 times higher than that of the LWF scheduler.

Figure 7 illustrates the increasing throughput obtained by relaxing the delay constraint (smaller α) which manifests as the density of the scheduled rates concentrating in the higher end of the range. Contrast the distribution of scheduled rates in the Maximum SNR scheduler and the LWF Scheduler. The former maximizes the gain from multiuser diversity and schedules users at higher rates. The CDF of the scheduled rates for the Maximum SNR scheduler is the product of the CDFs of the 16 users. Since there is a uniform probability of picking a user in any given slot, the distribution of the scheduled rate for the LWF scheduler is simply the distribution of the requested rates, $f_r(r)$ for any user as obtained from the Jakes model. In both figures, the distributions of the scheduled rate and vacation time for a value of $\alpha = 50$ lie between the two extremes.

D. Correlated Rates

In our analysis, we assume that the channel rates and therefore the requested rates for each user are independent of each other and identically distributed. Furthermore, the analysis also assumes that the requested rates are independent from one time slot to another. In our simulations, we also study the performance of the system when channel rates are correlated across time slots. With correlated fades, the channel remains in a *good* state or *bad* state across consecutive slots. As can be seen from Figure 8, which shows the CDF of the vacation time for $\alpha = 50$ and a doppler frequency, $f_d = 10$Hz, the probability of being scheduled at lower delays is higher when the requested rates are correlated. In this case the channel conditions dominate the metric. When the channel rates are correlated, there is also a higher probability of being scheduled when the vacation time is large. This may happen because (a) the user remains in a bad fade for a long duration, or (b) is pre-empted by other users with better channel conditions for a sustained duration. This is apparent from the tail of the vacation time distribution in Figure 8.

VII. CONCLUSIONS

The time-varying wireless channel capacity adds a new dimension to the problem of supporting broadband data services in cellular wireless networks. Implicit in the use of channel-state dependent scheduling algorithms are the questions of how these algorithms will address fairness and the provision of

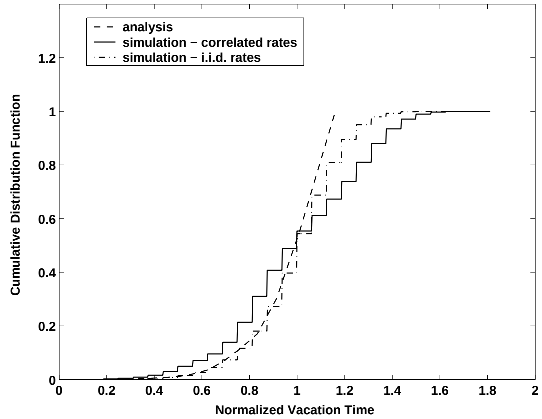

Fig. 8. Comparison of Vacaton Time CDFs for $\alpha = 50$ when channel rates are i.i.d. and correlated across time slots

QoS guarantees for a mix of traffic from delay-sensitive multimedia applications and non-real-time data traffic. The analytical results in this paper address the important issue of quantifying the QoS provided by a cellular wireless system. We completely characterize the distributions of the scheduled rates and delay in a general scheduler which realizes multiuser diversity gain with constraints on scheduling delay. In order to study this trade-off, we employ a general scheduling metric, $m = R + \alpha V$. The scheduler can be tuned to achieve the desired performance by varying the control parameter, α to balance the role of the channel rate, R or the delay, V in scheduling a user.

Fairness is an important performance criterion for a scheduler, given the large dynamic range of channel conditions in a cellular wireless system. The Proportional Fair (PF) [15] scheduler equitably shares system resources in a time-slotted system by ensuring all users get equal fractions of time-slots irrespective of their channel conditions. The scheduler metric proposed here can also be shown to be *resource fair* with a simple change in the metric. The modified metric for user i is

$$m_i(t) = (R_i(t) + \beta_i) + \alpha V_i(t) \tag{28}$$

where β_i can be chosen optimally to maximize the total scheduled rate while ensuring resource fairness without delay constraints ($\alpha = 0$). This result is proved in [12].

Consider the typical case of users at different nominal SNRs distributed throughout a cell. It would be reasonable to assume that statistical variations about the nominal SNR are identical for all users in the same cell. In such a scenario, any scheduler that exploits multiuser diversity gain alone will favor the users at higher nominal SNRs. Let $\gamma_i(t)$ and $\gamma_{NOM,i}$ represent the instantaneous SNR reported by the user and the nominal SNR of the user respectively. Since the rate requested by user i is proportional to $log(\gamma_i)$, setting $\beta_i = -log(\gamma_{NOM,i})$ in equation 28 is equivalent to a *Normalized* Maximum SNR Scheduler. As can be seen from Figure 9, this scheduler results in an equitable distribution of time slots among users with different nominal SNRs. For the case of 16 users with identical channel statistics about their nominal SNRs obtained from the Jakes channel model, we randomly distribute the nominal SNRs of the users in the range of 0dB to 8dB. While the Normalized Maximum SNR scheduler ensures a resource fraction that is roughly equal for all users, the Maximum SNR scheduler unfairly allocates a large fraction of time slots to users at the higher end of the range. In order to focus on the trade-off between throughput and delay alone, we assume that all users experience identical channel statistics in our analysis and therefore drop the fairness term, β_i, in the scheduling metric.

A natural question that would arise in the choice of a scheduler metric is whether the metric optimizes the scheduler performance with respect to some criterion of interest. The objective of this work is not so

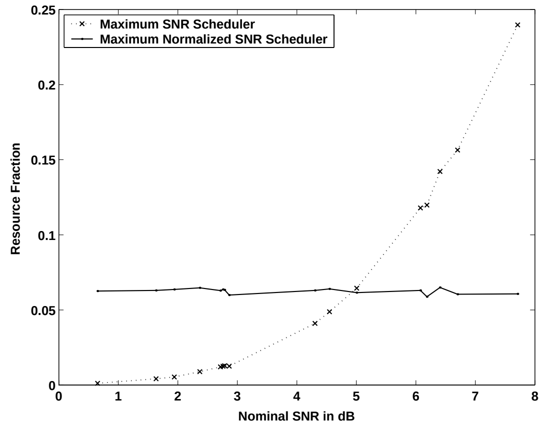

Fig. 9. Resource Fraction for 16 users at different nominal SNRs but identical channel statistics

much the design of an optimal scheduler as it is an analytical characterization of the trade-off between aggregate throughput and delay in a general opportunistic scheduler. The scheduler metric given by equation 1 can be configured to achieve a balance between multiuser diversity gain and delay constraints. The proposed scheduler lends itself well to statistical analysis while being sufficiently simple and versatile to be implemented in a real system. Our statistical analysis of user throughput and delay is validated by extensive simulations of a system architecture similar to a 1xEV-DO base station serving mobile users.

VIII. Acknowledgements

The first author is grateful to Tom Richardson, Sundeep Rangan and Murari Srinivasan, all of Flarion Technologies Inc., Bedminster, NJ, for their valuable comments which shaped this work.

Appendix
Computation of the Probability of Selecting a User in the Event of a Tie

This section details the computation of the probability of selecting a user in the event of a tie. Let $\mathcal{M}$ denote the set of unique values that the user metric, m_u, can take. Now consider a metric value, $m \in \mathcal{M}$. Suppose there are N_m users out of a total of N users who can take on this metric value. We define the event $\mathcal{E}_m$ as

$$\mathcal{E}_m = \text{two or more users with metric } m\text{, all other users with metric less than } m \tag{A-1}$$

This event may be represented as the union of mutually exclusive events as follows

$$\mathcal{E}_m = \bigcup_{k \geq 2} \mathcal{E}_{m,k} \tag{A-2}$$

where $\mathcal{E}_{m,k}$ denotes the event where exactly k users take on the metric m, with all other users having a lower metric. Since there are $\binom{N_m}{k}$ combinations of exactly k users from among N_m users, there are $\binom{N_m}{k}$ events that constitute the event $\mathcal{E}_{m,k}$. We denote this partition as

$$\mathcal{E}_{m,k} = \bigcup_{i} \mathcal{E}_{m,k}^{(i)} \tag{A-3}$$

Let $\mathcal{S}_{m,k}^{(i)}$ denote the particular set of k users which take on the metric value m to form the event $\mathcal{E}_{m,k}^{(i)}$. The probability of the event may be computed as

$$P(\mathcal{E}_{m,k}^{(i)}) = \prod_{u \in \mathcal{S}_{m,k}^{(i)}} P(R_u = m - \alpha V_u) \prod_{u' \in (\mathcal{S}_{m,k}^{(i)})^C} P(R_{u'} < m - \alpha V_{u'}) \tag{A-4}$$

A user $u \in \mathcal{S}_{m,k}^{(i)}$ is picked with uniform probability, $\frac{P(\mathcal{E}_{m,k}^{(i)})}{k}$. Finally, the probability of picking a user, u in the event that at least one other user has the same metric is

$$P_{tie-break}(u) = \sum_m P_{tie-break}(u,m), \quad \text{where} \tag{A-5}$$
$$P_{tie-break}(u,m) = \sum_{i,k} \frac{P(\mathcal{E}_{m,k}^{(i)})}{k}, \quad \forall i,k \text{ such that } u \in \mathcal{S}_{m,k}^{(i)}$$

References

[1] M. Grossglauser and D. Tse, *Mobility Increases the Capacity of Wireless Adhoc Networks*, IEEE Infocom '01, April 2001.
[2] R. Knopp and P. Humblet, *Information Capacity and Power Control in Single Cell Multiuser Communications*, IEEE ICC '95, Seattle WA, June 1995.
[3] D. N. C. Tse, *Optimal Power Allocation over Parallel Gaussian Channels*, in Proceedings, International Symposium on Information Theory, Ulm, Germany, 1997.
[4] T. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall, Upper Saddle River, NJ, 2002.
[5] 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, pp. 70-77,July 2000.
[6] TIA/EIA IS-856 , *CDMA 2000 : High Rate Packet Data Air Interface Specification*, Standard, November 2000.
[7] I. Bettesh and S. Shamai, *A Low Delay Algorithm for the Multiple Access Channel with Rayleigh Fading*, IEEE Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC '98, September 1998.
[8] I. Bettesh and S. Shamai, *Fading Channel : Information Theoretic and Communication Aspects*, IEEE Transactions on Information Theory, Vol 44, pp. 2619-2692, October 1998.
[9] M. Andrews, K. Kumaran, K. Ramanan, A.L. Stoylar, R. Vijayakumar, P. Whiting *CDMA Data QoS Scheduling on the Forward Link with Variable Channel Conditions*, Bell Labs Technical Report, April 2000.
[10] S. Shakkottai and A.L. Stoylar, *A Study of Scheduling Algorithms for a Mixture of Real and Non-Real Time Data in HDR*, Proc. 17th International Teletraffic Congress, September 2001.
[11] S. Shakkottai and A.L. Stoylar, *Scheduling for Multiple Flows Sharing Time-varying Channel : The Exponential Rule*, American Mathematical Society Translations, Series 2, A Volume in Memory of F. Karpelevich, Vol. 207, 2002.
[12] X. Liu, K.P. Chong and N.B. Shroff, *A Framework for Opportunistic Scheduling in Wireless Networks*, Computer Networks, vol. 41, no. 4, p. 451-474, 2003.
[13] R. Srinivasan and J. Baras, *Opportunistic Scheduling in Wireless Systems and TCP Performance*, Technical Report, TR 2002-48, I.S.R. University of Maryland, October 2002.
[14] M.C.Chan. R. Ramjee, *TCP/IP Performance over 3G Wireless Links with Rate and Delay Variation*, Mobicom, September 2002.
[15] D. N. C. Tse, *Opportunistic Communications: Smart Scheduling and Dumb Antennas*, presentation available at http://www.eecs.berkeley.edu/ dtse/intel128.pdf, January 2002.
[16] T.S.E. Ng, I.Stoica and H.Zhang, *Packet Fair Queueing Algorithms for Wireless Networks with Location-dependent Errors*, Proceedings, IEEE Infocom '98, pp. 1103-1111, March 1998.
[17] *Adapting Packet Fair Queueing Algorithms to Wireless Networks*, Proceedings, ACM Mobicom '98, October 1998.
[18] S. Lu, V. Bhargavan and R. Srikant, *Fair Scheduling in Wireless Packet Networks* IEEE ACM Transactions on Networking, vol 7, pp. 473-489, August 1999.
[19] P. Bhagwat, P. Bhattacharya, A. Krishna and S. Tripathi, *Enhancing Throughput Over Wireless LANs Using Channel State Dependent Packet Scheduling*, Proceedings, IEEE Infocom April 1996.
[20] C. Fragouli, V. Sivaraman and M. Srivastava, *Controlled Multimedia Wireless Link Sharing via Enhanced Class Based Queueing with Channel-state-dependent Packet Scheduling*, Proceedings of IEEE Infocom, pp.572-580, April 1998.
[21] T.S. Rappaport, A. Annamalai, R.M.Buehrer and W.H. Tranter, *Wireless Communications : Past Events and a Future Perspective*, IEEE Communications Magazine, vol 40, pp.148-161, May 2002.
[22] D. Wu and R. Negi, *Effective Capacity - A Wireless Channel Model for Support of Quality of Service*, To appear in IEEE Transactions on Wireless Networking 2003.
[23] W. C. Jakes, *Microwave Mobile Communications*, Piscataway, NJ : IEEE Press, 1993.