EURASIP Journal on
Wireless Communications and Networking
a SpringerOpen Journal

RESEARCH Open Access

Individual packet deadline delay constrained opportunistic scheduling for large multiuser systems

Muhammad Majid Butt[1,3*], Ralf R Müller[2,3] and Kimmo Kansanen[3]

Abstract

This work addresses opportunistic distributed multiuser scheduling in the presence of a fixed packet deadline delay constraint. A threshold-based scheduling scheme is proposed which uses the instantaneous channel gain and buffering time of the individual packets to schedule a group of users simultaneously in order to minimize the average system energy consumption while fulfilling the deadline delay constraint for every packet. The multiuser environment is modeled as a continuum of interference such that the optimization can be performed for each buffered packet separately by using a Markov chain where the states represent the waiting time of each buffered packet. We analyze the proposed scheme in the large user limit and demonstrate the delay-energy trade-off exhibited by the scheme. We show that the multiuser scheduling can be broken into a packet-based scheduling problem in the large user limit and the packet scheduling decisions are independent of the deadline delay distribution of the packets.

Keywords: Opportunistic scheduling; Large system analysis; Delay constrained data; Energy efficient communications

1 Introduction

We consider a wireless communication system with K users and a single central base station. Each user is subject to both time-varying frequency-selective fading and position-dependent path loss. This setting was addressed before in, e.g., [1], where proportional fair scheduling was compared to hard fair scheduling. While proportional fair scheduler [2] does not guarantee any upper bound on the delay of a data packet, hard fair scheduling enforces that each data packet is scheduled instantaneously. Packet delay can further be classified into average tolerable delay and maximum tolerable delay. This work focuses on the later definition of delay which is also called packet deadline.

In a practical system, information will be outdated after a certain delay time has passed and scheduling an outdated packet will be obsolete. There are two proper approaches to deal with the fact that packets become outdated: Either to drop them if they have not been scheduled in time or to force their transmission if they are reaching their deadline. Which way is more appropriate to go depends on the particular application, i.e. on the potential damage caused by a lost packet. In both cases, there is a trade-off between delay, throughput and power consumption.

Reference [3] deals with the trade-offs between average delay and average power. Reference [4] uses multiuser diversity to provide statistical quality of service (QoS) in terms of data rate, delay bound, and delay bound violation probability. In [5], an exact solution for the average packet delay under the optimal offline scheduler is presented when an asymmetry property of packet inter-arrival times and packet inter-transmission times holds. Online scheduling algorithms that assume no future packet arrival information are discussed as well. Their performances are comparable to those of the offline schedulers which assume identically and independently distributed inter-arrival times. The results of [3] have been extended to the multiuser context in [6]. It is found that to achieve an average power within the $\mathcal{O}(1/V)$ of the minimum power required for network stability, there must

*Correspondence: majid.butt@ieee.org
[1] Department of Computer Engineering, Qatar University, Doha, Qatar
[3] Institute of Electronics and Telecommunications, Norwegian University of Science and Technology, Trondheim, Norway
Full list of author information is available at the end of the article

be an average queuing delay greater or equal to $\Omega(\sqrt{V})$, where $V > 0$ is a control parameter.

In [7], the authors consider the energy minimization problem for packet deadline-constrained applications. The channel of each user is discretized to one of a finite number of states. They consider two cases of rate-power curves. For both the cases, they obtain dynamic programming-based optimal solutions. When the rate-power relation is linear, they obtain a threshold-based scheduler which follows the optimal stopping theory formulation in [8]. For the case of a convex rate-power curve, a heuristic algorithm is proposed which gives a solution quite close to the optimal. A similar approach is applied in [9] where the authors consider the same problem for a point-to-point network. They consider a packet of B bits which has to be transmitted within the hard deadline of N time slots. During the transmission of the packet, no other packets are scheduled. The authors obtain close form expressions for the optimal policy only for the case $N = 2$ using dynamic programming. For $N > 2$, the optimal policy is numerically determined. It should be noted that the optimal solution is obtained only when either the rate-power curve is linear [7] or scheduling of a single packet is considered [9] following the framework of optimal stopping theory.

The problems of finding optimum solutions and the need for dynamic programming result from the interdependence of the users' scheduling decisions. However, as the number of users becomes large, the instantaneous effect of the other users converges to its statistical average and optimum scheduling decisions can be made by each user individually without considering the fading states and queue lengths of the other users. In this context, this principle was first reported in [10]. It runs under various names in literature, e.g., large-system limit, mean-field approach, self-averaging, etc. For a more general discussion on the range of its applicability, see, e.g., [11,12]. The many-user limit was applied in [13] and an algorithm called opportunistic superpositioning (OSP) was proposed to provide all users their desired average data rates while guaranteeing a certain average delay. The average delay of the users is inversely proportional to the scheduling probability and the scheduling threshold is used to control the delay. In the many-user limit, it is shown analytically that the required power can be made arbitrarily small at the expense of increased average delay.

In contrast to [13] and most other works discussed above, this paper addresses a system with a strict packet deadline (and not average) delay constraint. The packet deadline delay varies from packet to packet. The aim is to minimize the system energy while obeying the packet deadline delay constraint for each arriving packet. We first address the *many-user* limit where scheduling decisions can be taken based on each user's own queue without loss of optimality. In this context, scheduling is not restricted to schedule one user at a time, but to schedule a finite fraction of the users, which experience favorable channel conditions and/or whose packets are close to their deadline, simultaneously. Though these users interfere with each other, they can be separated by means of superposition coding. Their effects onto each other decouple in the many-user limit and we can reformulate the multiuser scheduling problem as an equivalent single-user scheduling problem following the lines of thought in [14]. To the best of our knowledge, packet deadline-based scheduling has not been addressed in multiuser settings before. We apply the scheduling strategy which we find optimum in the many-user limit to the finite-user case and show that it, though suboptimum there, performs very well. We generalize the approach in [15], where an identical deadline is assumed for all the arriving packets and the simplified multiuser scheduler is limited to the policy of either scheduling all the buffered packets (simultaneously) or waiting for the next time slot. In this work, we provide a complete mathematical framework for the energy optimal packet-based scheduling and analyze the proposed scheme using Markov chain in the many-user limit. We show analytically that the scheduling decisions are independent of the deadline distribution but system energy depends on the deadline distribution. We discuss stochastic optimization techniques for optimization and show that the complexity in computing the thresholds remains acceptable.

The remainder of this paper is organized as follows: Section 2 describes the system model and Section 3 addresses the many-user considerations used in this work. The proposed multiuser scheduling scheme is presented in Section 4. The steady-state analysis of the queue is discussed in Section 5. We discuss the optimization procedure for the proposed scheme in Section 6. In Section 7, implementation issues of the proposed scheme are considered while numerical results are presented in Section 8. Section 9 concludes with the main results and contributions of this paper.

2 System model

We consider a multiple-access system with K users randomly placed within a certain geographical area. Each user is provided a certain fraction of the resources available to the system. We consider a time-slotted system. Arrivals occur at the start of a time slot and are queued in a finite buffer before transmission. Scheduling is performed at the end of a time slot taking into account the new arrivals within the current time slot. We consider an uplink (reverse link) scenario but the results can be generalized to a downlink (forward link) scenario in a straightforward manner using the multiple-access broadcast duality of the Gaussian channel [16] and the fact that scheduling decisions decouple in the many-user limit.

2.1 Channel model

The fading environment of the multi-access system is described as follows. We model the frequency-selective short-term fading by a multi-band channel with independent Rayleigh fading within each band. Each user k experiences a channel gain $g_k(t)$ in slot t. The channel gain $g_k(t)$ is the product of path gain s_k and short-term fading $f_k(t)$ i.e. $g_k(t) = s_k f_k(t)$. Path loss and short-term fading are assumed to be independent. The path gain is a function of the distance between the transmitter and the receiver and we assume it not to change within the time-scales considered in this work. Short-term fading depends on the scattering environment and occurs when the coherence time of the channel is shorter than the delay requirement of the application. Short-term fading changes from slot to slot for every user and is independent and identically distributed across both users and slots but remains constant within each single transmission. This model is often referred to as block fading. For a multi-band system with M channels, short-term fading over the best channel is represented by $f_k(t) = \max\{f_k^{(1)}(t), f_k^{(2)}(t), \ldots, f_k^{(M)}(t)\}$.

$E_k^R(t)$ and $E_k(t)$ respectively represent the received and the transmitted energy per symbol of each user k such that

$$E_k^R(t) = g_k(t)E_k(t). \tag{1}$$

Note that the distribution of $g_k(t)$ differs from user to user. Let N_0 denote the noise power spectral density. The channel state information is assumed to be known at both the transmitter and the receiver side. This can be accomplished by channel estimation on the opposite link (downlink) in time-division duplex systems or by communication of explicit side information within the coherence time of the channel.

2.2 Physical layer communication

It is mandatory to allow multiple users to be scheduled simultaneously in the same time slot and in the same frequency band. Otherwise, a packet deadline of a finite number of time slots could not be met without allowing non zero dropping probability[a], as the number of packets that have reached the deadline could exceed the number of available frequency bands.

In our settings, there is no limit on the number of users scheduled simultaneously thanks to many-user considerations (discussed in the following); and a theoretical framework with zero outage probability is considered without loss of generality.

The simultaneously scheduled users are separated by superposition coding. Let $\mathcal{K}_m$ be the index set of users to be scheduled in frequency band m. Let $\psi_1^{(m)} \cdots, \psi_k^{(m)}, \ldots, \psi_{|\mathcal{K}_m|}^{(m)}$ be a permutation of the scheduled user indices for frequency band m that sorts the channel gains in increasing order, i.e. $g_{\psi_1^{(m)}} \le \cdots \le g_{\psi_k^{(m)}} \le \cdots \le g_{\psi_{|\mathcal{K}_m|}^{(m)}}$. Then, the energy per symbol of user $\psi_k^{(m)}$ with rate $R_{\psi_k^{(m)}}$, as scheduled by the scheduler to guarantee error free communication, is given by [1,17]

$$E_{\psi_k^{(m)}} = \frac{N_0}{g_{\psi_k^{(m)}}}\left[2^{R_{\psi_k^{(m)}}} - 1\right] 2^{\sum_{i=1}^{k-1} R_{\psi_i^{(m)}}}. \tag{2}$$

This energy assignment results in the minimum total transmit energy per symbol for the scheduled users. On the receiver side, the data from the user with the worst channel is decoded first, treating the signals from all other users as noise. The data from the current user is decoded after decoding the data from the previous users whose signals have been subtracted from the received signal. All users are decoded by repeating this step *successively*. This is the well-known successive interference cancelation (SIC). Collisions between simultaneous transmissions are avoided because in a multiuser environment, superposition coding and successive decoding ensure that data from multiple users are decoded successfully without errors on the receiver side[b].

2.3 Queuing model

At each time slot, none, one, or several packets arrive to the queue of each user. In general, an arriving packet is characterized by two parameters: its size and its deadline. Formally, the deadline is defined as the number of time slots available *at the arrival time* of the packet in the buffer before it has to be scheduled irrespective of channel conditions.

Without loss of generality, all packets are assumed to have a unit size. Note that larger packets can be modeled as being composed of multiple virtual packets of unit size. The deadlines of the packets are assumed to be finite and positive but arbitrary, otherwise. We model the arrival process by the probabilities p_i that give the probability that an arriving virtual packet has deadline τ_i with $i \in \{1 \ldots N\}$. The maximum size of the user's buffer N is a system parameter and is given by the *maximum* of the deadlines $\tau_i \forall i$ of all the packets in the system.

For each packet in a queue, a decision is made whether it is scheduled at the present time slot or not. There is no limit on the maximum number of (virtual) packets in the queue. The system considered in our settings is entirely driven by the demands of the users. Each user's demand on rate and delay has to be met by the system. Packet drops or outage are strictly prohibited. Since data rate and energy can be freely exchanged against each other, the users' demands can always be met with sufficient use of energy. Though, the higher the demands of the users, the more energy the system will consume. However, the system has certain degrees of freedom to reduce the energy consumption: It can decide when a certain packet is being transmitted within the time left to its delay deadline. The system can decide whether to split packets into subpackets. These sub-packets can then be either transmitted simultaneously, transmitted at different times, or combinations of the these two options can be used. Furthermore, the system can decide which frequency bands to use for which user's packets at which time. It may seem infeasible to build a system that can find the optimum strategy to schedule each packet at the right time. However, we will make two idealized assumptions that allow us to characterize the structure of the optimum scheduling policy up to a few parameters that can be optimized numerically. First, we assume that there exists a coding strategy that achieves the capacity region of the Gaussian multiple-access channel. State-of-the-art coding strategies for the Gaussian multiple-access channel are indeed very close to the capacity region [18,19]. Second, we assume that the number of users and the available radio spectrum grow asymptotically large, with the ratio of the number of users to radio spectrum being constant. This assumption is a good approximation for a system, where the individual user's data rate is much lower than the total data rate of the system [20].

3 Large-system considerations

Consider the average energy per symbol and the total rate of all users in all bands

$$E_{\mathrm{s}} = \frac{1}{K}\sum_{k=1}^{K} E_k \tag{3}$$

$$R = \sum_{k=1}^{K} R_k, \tag{4}$$

respectively, and denote the average energy per bit as

$$E_{\mathrm{b}} = E_{\mathrm{s}}/R. \tag{5}$$

Total rate and energy per bit are system parameters that must be finite and positive irrespective of the system size. Due to (3) and (4) and many-user considerations when $K \to \infty$,

$$R_k = \mathrm{O}(1/K) \tag{6}$$

for all users. Note that due to (6), E_k, the energy per symbol for user k in (2) is a linear function of R_k, the rate of user k in the many-user limit. Remarkably, this simplicity is inherent by the system (similar to multiuser diversity) due to the presence of large number of users in the system and we quantify in Section 8 that a few hundred users are enough to achieve the asymptotic results. The linearity of the energy per symbol greatly simplifies the scheduling decisions. Based on this, we have

Lemma 1. *In the many-user limit, scheduling decisions in the queue of a user k can be made on a packet-by-packet basis without loss of optimality. Furthermore, the optimal scheduling decision does not depend on the properties of the other packets in the same queue.*

The lemma implies that we cannot save energy by scheduling only some of several packets of a user with the same number of remaining time slots before deadline, as the energy costs of the packets are additive due to (6) (and not exponential as appears in (2)). Thus, independence of scheduling decisions for every packet remains optimal.

Additionally, we can decouple scheduling decisions among different users based on many-user assumptions and our discussion in Section 1 [10,11,15].

Lemma 2. *In the many-user limit, scheduling decisions can be made on a user-by-user basis without loss of optimality. Furthermore, the optimal scheduling decisions for a queue of a user do not depend on the properties of the queues of the other users.*

By applying many-user assumptions, Lemma 2 breaks the joint multiuser scheduling problem into an equivalent single user scheduling problem [15] while Lemma 1 decomposes the problem further into individual packet deadline-dependent scheduling.

3.1 State space model

In the following, we develop a Markov decision process (MDP)-based model for the scheduling of deadline-dependent packets. We define the *state* of the MDP as the number of time slots remaining before a packet (virtual user) has to be scheduled irrespective of the fading conditions. The definition of the state appears to be very similar to the definition of the deadline in Section 2.3. However, the deadline is a system parameter associated with a packet at the time of arrival and is fixed. The state of a packet varies over the period of time it spends in the buffer. At the start of the MDP process, the state equals the deadline. In each subsequent time slot, if the packet is not scheduled, it decreases by one until it reaches one. The system parameter N defined in Section 2.3 determines the size of the Markov chain.

With a modest amount of foresight, let us decompose all the packets queued with each user into N virtual users: one virtual user for each state. Note that all packets in the buffer of a virtual user have the same state and there is no limit on the size of the buffer for a virtual user. Every newly arriving packet with deadline τ_i is put into the buffer of the ith virtual user. The schematic diagram for a two-dimensional buffer for a user's buffer has been shown in Figure 1.

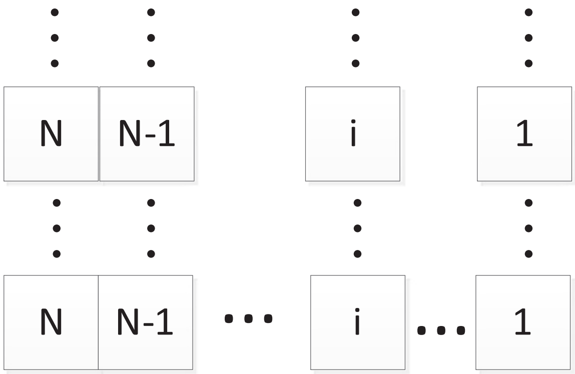

Figure 1 Schematic diagram for the buffer of a user which consists of N virtual buffers (users) of infinite size.

When being scheduled, the virtual buffer is emptied at once. The scheduling decision of the virtual buffer is explained in the next section. The rate of a virtual user correlates with its fading. The better a user's channel, the higher the probability that the rate is non-zero. Let us now introduce decision variables $d_{k,i}$ for all virtual users (k, i) that indicate whether the packets of the virtual user (k, i) are scheduled. Then, conditioned on $d_{k,i}, k = 1..K, i = 1..N$, the rates of the virtual users are independent of their fading. Due to this conditional independence, we have in the many-user limit $(K \to \infty)$ [1]

$$\frac{E_\mathrm{b}}{N_0} = \log(2) \int_0^\infty \frac{2^{R\,\mathrm{P}_{g|d=1}(x)}}{x}\, \mathrm{dP}_{g|d=1}(x) \tag{7}$$

where $P_{g|d=1}(x)$ denotes the distribution of the fading of the scheduled virtual users. Remarkably, the rates of the users affect (7) only *via* its total sum R.

4 Threshold based scheduling scheme

Scheduling is a decision process. We adopt a fading threshold-based policy which quantizes the fading vector into a finite number of intervals. These intervals depend on the state of the packet and the fading distribution. We introduce (quantized fading states) thresholds to determine whether a packet (virtual buffer) with a state i is scheduled or not. These thresholds may depend on all system variables in general. However, in the many-user limit, they will depend only on each user's own parameters, i.e. fading and state.

Definition 1 (Transmission threshold)**.** *A transmission threshold κ_i is defined as the minimum short-term fading value allowing for scheduling a packet (virtual user) with state i.*

Note that scheduling decisions only depend on the short-term fading. This is easily proven by contradiction. Imagine scheduling decisions would depend on the path loss. This would not lead to unstable queues due to the hard deadline constraint. However, it would cause a greater average delay of users with worse path loss compared to users with better path loss. In fact, the path loss would be reflected as a bias in the average queuing time of packets and such a bias reduces the dynamics of the scheduling process. This is clearly an adverse effect.

Next, we state a few fundamental properties of these transmission thresholds.

Property 1. *There is no minimum fading value required to schedule a packet that has reached its deadline, i.e.*

$$\kappa_0 = 0. \tag{8}$$

This ensures that the hard deadline is kept regardless of the channel quality[c].

Property 2. *The closer the packets are to the deadline, the more likely they are to be scheduled, i.e.*

$$\kappa_{i+1} > \kappa_i \qquad \forall i. \tag{9}$$

This is evident from the construction of the problem that the probability of scheduling of a packet must be increased as it comes close to the deadline which is achieved by reducing the channel-dependent threshold with decreasing state i.

In order to ease notation, we introduce an additional state $N + 1$. We model the packet being in that state when it is not in the queue, i.e. before it has arrived and after it has been scheduled.

The probabilities of the state transitions $T_{N+1\to i}, \forall i$ model the statistics of the random arrival process

$$\alpha_{N+1\to i} = \Pr(T_{N+1\to i}) = p_i \qquad \forall 1 \le i \le N \tag{10}$$

where p_i denotes the probability that an arriving packet has deadline τ_i, cf. Section 2.3. A packet with deadline $\tau_i < \tau_N$ is inserted directly into state i and treated as a packet that arrived in the buffer with deadline τ_N but has not been scheduled for $N - i$ time slots. This reduces degrees of freedom available for the packet and results in high energy cost.

The probabilities of the state transitions $T_{i\to N+1}, \forall i$ are determined by the transmission thresholds as follows

$$\alpha_{i\to N+1} = \Pr(T_{i\to N+1}) = \Pr(f \ge \kappa_i) \qquad \forall 1 \le i \le N, \tag{11}$$

where f denotes the short-term fading as explained in Section 2.1. We remove subscript k as all the users have identical fading distribution.

$\alpha_{i\to N+1}$ are the probabilities of being scheduled, while

$$\begin{aligned}\alpha_{i\to i-1} &= \Pr(T_{i\to i-1}) = \Pr(f < \kappa_i)\\ &= 1-\Pr(T_{i\to N+1}) \qquad \forall 1 < i \le N\end{aligned} \tag{12}$$

are the probabilities of being not scheduled. All other state transitions are impossible. A state transition diagram is depicted in Figure 2.

5 Distribution of packet deadlines

Our modeling of the problem ensures that the scheduling decisions and the thresholds are independent of the deadline distribution of the packets. However, the average system energy expenditure depends on the deadline distribution. In the limiting case $K \to \infty$, the empirical average of the arrival rate converges uniformly to its expectation $\lambda = R/K$. However, the buffer occupancy of the scheduled states is not uniform and depends on the deadline and fading distributions. A variable buffer occupancy model helps us in understanding the energy behavior of the system as a function of deadline distribution of the arriving packets and the fading distribution. For example, a large value of p_1 means more degree of freedom in scheduling and small energy expenditure while large value of p_N implies strict latency requirements and large energy expenditure.

Let us consider the buffer occupancies for the different states in the limiting case $K \to \infty$: The average number of packets getting into state i must equal the average number of packets getting out of that state. Thus, we have for $i < N$

$$p_i\lambda + \alpha_{i+1\to i}L_{i+1} = L_i \tag{13}$$

with

$$L_N = p_N\lambda \tag{14}$$

where L_i is the average number of virtual packets in state i. The steady-state probability that a packet in the queue is in state i is thus given by

$$\begin{aligned}\pi_i &= \frac{L_i}{L} &(15)\\ &= \frac{p_i\lambda}{L} + \alpha_{i+1\to i}\pi_{i+1} &(16)\\ &= \frac{\lambda}{L}\sum_{n=i}^{N} p_n \prod_{m=i+1}^{n} \alpha_{m\to m-1} &(17)\end{aligned}$$

where we define $\pi_{N+1} = 0$ and

$$L = \sum_{i=1}^{N} L_i \tag{18}$$

for notational convenience[d]. With these steady-state probabilities, the distribution of the fading of the scheduled users can be calculated. Furthermore, we note that

$$\frac{L}{\lambda} \tag{19}$$

is the ratio of packets in the queue to the number of packets arriving. This is the average delay of the system.

6 Threshold optimization

Next, we would like to optimize the transmission thresholds. Our objective is to minimize the average transmitted energy per symbol given in (7) for the constraint that every packet is scheduled before reaching the deadline. Energy per symbol depends solely on the channel distribution $\mathrm{P}_{g|d=1}(.)$ of the scheduled virtual users (SVUs). The channel distribution of SVUs is a function of transmission thresholds or transition probabilities (interchangeably) and computed in the following based on the MDP model developed in the previous section.

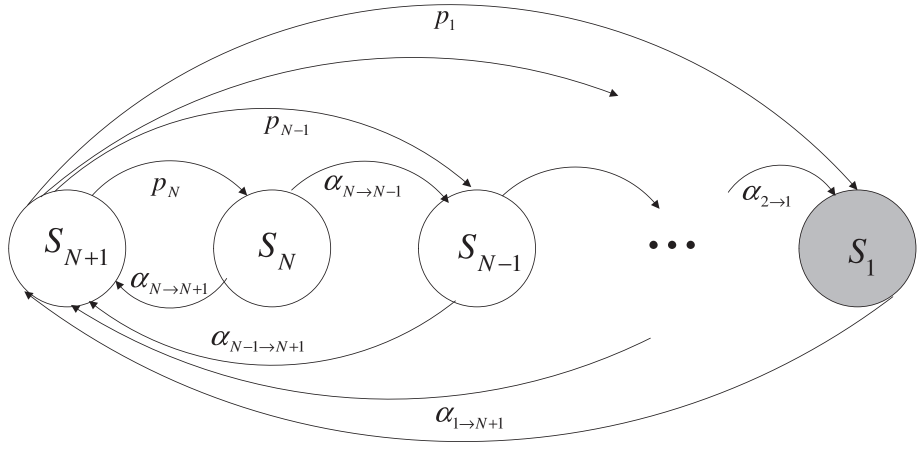

Figure 2 State diagram of proposed scheduling scheme where every state $i \neq N+1$ represents distance of buffered packet from deadline.

Equivalently, we formulate the optimization problem as,

$$\min_{\vec{\alpha}\in\Omega} \frac{E_\mathrm{b}}{N_0} \tag{20}$$

$$\text{subject to}: \begin{cases} \mathcal{C}_1 : 0 \le \alpha_{i\to N+1} \le 1 \quad 1 \le i \le N \\ \mathcal{C}_2 : \sum_{j=1}^{N+1} \alpha_{i\to j} = 1 \quad 1 \le i \le N, 1 \le j \le N+1 \\ \mathcal{C}_3 : \alpha_{1\to N+1} = 1 \end{cases} \tag{21}$$

where $\vec{\alpha} = [\alpha_{N\to N+1} \cdots \alpha_{1\to N+1}]$ and Ω defines the possible vector space for $\vec{\alpha}$ with $\vec{\alpha}$ containing all the transition probabilities representing scheduling of a packet (decision variables). $\mathcal{C}_1$ and $\mathcal{C}_2$ follow the properties of homogenous Markov chain while $\mathcal{C}_3$ results from Property 1 of transmission thresholds. For the optimized $\vec{\alpha}^*$, the corresponding transmission threshold vector $\vec{\kappa}^* = [\kappa_N^* \ldots \kappa_1^*]$ can be computed using (11) and vice versa.

To compute the solution of the optimization problem, we need to express the probability distribution of the fading of SVUs $\mathrm{P}_{g|d=1}(\cdot)$ in (7) in terms of $\vec{\kappa}$.

Using Bayes' law, the probability density function (pdf) of the short-term fading of the SVUs is given by

$$\mathrm{p}_{f|d=1}(y) = \mathrm{p}_f(y)\,\frac{\Pr(d=1|f=y)}{\Pr(d=1)} \tag{22}$$

$$= \frac{\sum\limits_{i=1}^{N} \pi_i\, 1(y \ge \kappa_i)\, \mathrm{p}_f(y)}{\int \sum\limits_{i=1}^{N} \pi_i\, 1(y \ge \kappa_i)\, \mathrm{dP}_f(y)} \tag{23}$$

$$= \frac{\sum\limits_{i=1}^{N} \pi_i\, 1(y \ge \kappa_i)\, \mathrm{p}_f(y)}{1 - \sum\limits_{i=1}^{N} \pi_i\, \mathrm{P}_f(\kappa_i)} \tag{24}$$

where the denominator results from integration by parts and $1(\cdot)$ is 1 if the argument is true and 0 if the argument is false. Using integration by parts once more, we find the CDF as

$$\mathrm{P}_{f|d=1}(y) = \frac{\sum\limits_{i=1}^{N} \pi_i 1(y \ge \kappa_i)\left[\mathrm{P}_f(y) - \mathrm{P}_f(\kappa_i)\right]}{1 - \sum\limits_{i=1}^{N} \pi_i\, \mathrm{P}_f(\kappa_i)}. \tag{25}$$

Using standard methods for calculating the distribution of the product of two independent random variables, $\mathrm{P}_{g|d=1}(y)$ is calculated in the Appendix from (24) and the CDF of the path loss.

The energy in (7) is not a convex function of the transmission thresholds. In the following, we discuss two heuristic optimization techniques to compute transmission thresholds.

6.1 Optimization by simulated annealing

We choose to use the simulated annealing (SA) algorithm to optimize the energy function for the transmission thresholds that result in the minimum energy for a given deadline delay parameter. The simulated annealing algorithm was proposed in [21] and [22] separately. It uses ideas from statistical mechanics to solve combinational problems. It is believed to provide near-optimal solutions (even optimal) in many combinatorial problems.

The main components of the simulated annealing algorithm are described briefly here.

1. Objective function
   In this work, the objective function is the system energy as given in (7).
2. Description of the configuration of the system
   It is essential to provide a clear description of the configuration of the system. In our case, the vector $\vec{\alpha}$ is the parameter which represents the configuration of the system at a particular instant. The transmission thresholds are related to the transition probabilities for a given deadline and short-term fading.
3. A random generator for the new configuration
   At the start of the algorithm, any configuration can be provided. In the next step, there must be a suitable method to provide a random change in the configuration. In this work, transition probability vector $\vec{\alpha}$ is varied in each step to provide a new configuration to evaluate (7).
4. A cooling temperature schedule
   The system is 'heated' at high temperature T at the start of the algorithm. Afterwards, the temperature is decreased slowly up to the point where the system 'freezes'. The term heating and cooling originate in statistical thermodynamics where freezing of the system represents a situation where the system reaches a near-optimal solution and no further state[e] transitions occur for further decrease of the temperature parameter. The cooling schedule depends on the specific problem and can be developed after certain experiments. In our simulations, we tested both Boltzmann annealing (BA) and fast annealing (FA) temperature cooling schedules which have been proven to provide global minimum solutions for a wide range of problems [23,24]. In FA, it is sufficient to decrease the temperature linearly in each step q such that,

   $$T_q = \frac{T_0}{q+1} \tag{26}$$

   where T_0 is a suitable starting temperature. Similarly in BA, global minima can be found sufficiently (in many problems) if temperature decreases logarithmically such that,

$$T_q = \frac{T_0}{\ln(q+1)} \tag{27}$$

5. Acceptance probability
   Any new configuration in SA is accepted if it results in a lower system energy with probability 1. A change in energy in each step is denoted by ΔE. Any new state is accepted with probability $\Delta E/T$ if it results in a higher energy state and it is referred to as *muting*. Muting occurs frequently at the start of the algorithm and vanishes to happen as the temperature T approaches zero.

Using the SA algorithm, an optimal vector $\vec{\alpha}*$ is obtained for a given N. The muting step makes it likely that local minima are avoided in the optimization process by moving into higher energy solutions with some temperature-dependent probability. Flow chart for SA algorithm for computation of thresholds has been shown in Figure 3. Numerical results for the

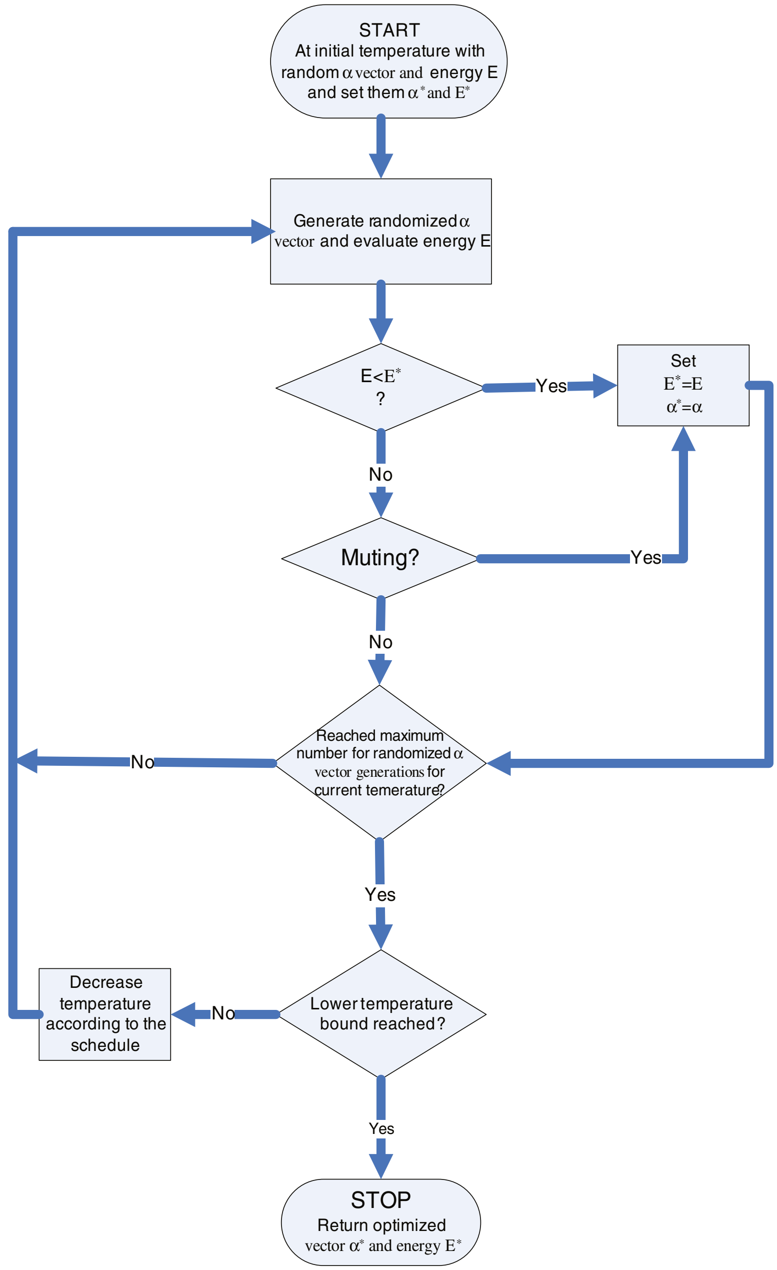
START
At initial temperature with random α vector and energy E and set them α* and E*
Generate randomized α vector and evaluate energy E
E<E* ?
Yes
Set E*=E α*=α
No
Muting?
Yes
No
Reached maximum number for randomized α vector generations for current temerature?
No
Yes
Lower temperature bound reached?
No
Decrease temperature according to the schedule
Yes
STOP
Return optimized vector α* and energy E*

Figure 3 Flow chart for implementation of SA for optimization of thresholds.

optimization process using SA algorithm are discussed in Section 8.

6.2 Optimization by recursion

This approach stems from the dynamic programming area where recursive optimization is used to compute the thresholds for problems belonging to optimal stopping theory. The optimized transmission threshold vector is found using a recursive procedure explained in the following:

1. Start the optimization procedure for $N = 2$ such that the optimization is a scalar problem and we only need to find the threshold κ_N since $\kappa_1 = 0$.
2. Given the optimized threshold vector[f] for N, i.e. $\vec{\kappa}^*(N) = [\kappa_N^*(N), \kappa_{N-1}^*(N), \ldots, \kappa_2^*(N), 0]$, we find the threshold vector for the deadline $N + 1$ by the heuristic postulate $\vec{\kappa}^*(N + 1) = [\kappa_N^*(N + 1), \vec{\kappa}^*(N)]$ and optimize over $\kappa_N^*(N + 1)$. Again, this is a scalar optimization problem.

The postulate $\vec{\kappa}^*(N + 1) = [\kappa_N^*(N + 1), \vec{\kappa}^*(N)]$ helps to reduce complexity for computation of thresholds significantly. In SA, computation complexity for computing thresholds is $\mathcal{O}(N)$ while recursive method requires just one additional threshold as $N - 1$ thresholds are known. We show in Section 8 that the results produced by both of the heuristic algorithms are indistinguishable.

7 Implementation considerations

The proposed scheme solves the optimization problem offline as a function of the channel statistics and state for each buffered packet. The offline optimization task can be performed locally by the users and needs no centralized control since it only involves the fading statistics, but not the fading realizations. However, a centralized optimization would save complexity, since the outcome of the optimization is identical for all users. Similarly, the scheduling decisions can fully be taken by each user individually. However, the powers required by the users to transmit their packets depend on the ordering of the successive decoding. For a finite user system, it is not possible for the users to get the exact knowledge of the required transmit power to provide the rate. Therefore, the users need to transmit with a power margin. The average excess power of the users should vanish in the many-user limit such that the system obeys (7). This does not happen if successive decoding is used. However, joint decoding does not suffer from this problem as all the users are decoded at the same time (without a specific order). Thus, for successive decoding, there is a need for a centralized assignment of the transmit powers. If the number of scheduled users, however, is very large and joint decoding is employed, the users can calculate their transmit powers individually by closely approximating the empirical fading and rate distributions of the other scheduled users by their statistical averages following the ideas of [10]. With the application of joint decoding, the proposed scheme has the potential to be implemented in a distributed manner.

The simplicity in making the scheduling decisions based on comparing the offline computed thresholds with channel conditions is well-suited to delay sensitive applications and power-limited devices. By using the parameter τ_i and the deadline distribution, we can control the energy-delay trade-off. A large value of τ_i (and corresponding large p_i) implies that the application data is more delay tolerant and the energy consumption will be closer to the energy consumption of the schemes without deadline delay guarantees.

8 Numerical results

We consider a multiple-access channel with M bands and assume that the short-term fading of the channels is statistically independent. Every user senses M channels and selects its best channel as the candidate for transmission. Therefore, a specific user is scheduled if its *best* channel is greater than the transmission threshold. This is the optimal multi-band allocation for the *hard fairness* asymptotic case [1]. Spectral efficiency is normalized by M to get spectral efficiency per channel C. We consider a system where users are placed uniformly at random in a cell except for a forbidden region around the access point of radius $\delta = 0.01$. The path loss is monomial with exponent 2. All users experience fast fading with exponential power distribution with unit mean on each of the M channels. The details of the path loss model can be found in the Appendix.

For all numerical results, the SA algorithm used 50 random configurations per temperature iteration. For a single-channel scheduler and a target spectral efficiency of $C = 0.5$ bits/s/Hz, the thresholds optimized with SA are shown in Table 1. The corresponding recursively computed threshold values are shown in Table 2. We find insignificant energy differences in the results computed by the two heuristic algorithms. This is easily understood due to the minor (and no) differences in the threshold assignments resulting from the two methods. Clearly, the recursive algorithm is preferable due to its significantly lower complexity.

Figures 4 and 5 show the statistics for the SA algorithm with the FA temperature cooling schedule. As explained in

Table 1 Thresholds computed *via* SA

N	κ_4	κ_3	κ_2	κ_1	$(E_b/N_0)_{sys}$
2	-	-	0.24	0	−1.42 dB
3	-	0.54	0.23	0	−3.06 dB
4	0.76	0.59	0.22	0	−4.07 dB

Table 2 Recursively computed thresholds

N	κ_4	κ_3	κ_2	κ_1	$(E_b/N_0)_{sys}$
2	-	-	0.24	0	−1.42 dB
3	-	0.52	0.24	0	−3.05 dB
4	0.75	0.52	0.24	0	−4.08 dB

Section 6.1, mutations occur with 100% probability at the start and then their frequency decreases with every iteration. Similarly, energy updates are more frequent at the start. Once the system finds the minimum energy solution, no more updates occur in spite of the occurrence of muting. It should be noted that statistics can differ a bit for different cooling schedules (like BA) and different configuration schedules, but the final results remain unaffected.

Figure 6 demonstrates the average system energy for a delay limited system. As the deadline of transmission for the packets increases, the average system energy decreases. Obviously, a trade-off between delay tolerance and energy consumption occurs which is more noticeable at smaller spectral efficiencies. Moreover, savings in system energy are more pronounced when N varies from 1 to 2 as compared to the case when N varies from 4 to 5. This effect is similar to time diversity where performance improvement is more pronounced at the addition of a few initial degrees of diversity.

Figure 7 demonstrates the effect of frequency diversity on the proposed scheduler for $N = 2$. A unique set of thresholds need to be optimized for a specific number of channels as optimal thresholds change with the number of channels. As explained in the system model, a user selects his best channel as a candidate for transmission and makes scheduling decisions by comparing the best channel with the thresholds. If there are more channels available for selection, the best channel (maximum value) improves with the number of the channels which in turn helps to reduce energy expenditure for the user. Thus, the number of channels provides an additional degree of freedom to further improve the energy consumption of the system.

Figure 8 shows the effect of finite number of users on the scheduler for a system with $M = 10$ for both constant and random arrivals. The results are obtained by varying the number of users in (2) which is a finite user approximation of the asymptotic expression in (7). For the numerical results, 250 simulations with different fading values have been performed for a single path loss. For a fixed number of users and iterations, we compute and compare the variance of the system energy for the cases of constant and random arrivals. We used a Bernoulli random arrival process in this example. The variance of the system energy for both constant and random arrivals decreases fastly as the number of users increases. For the same number of users, the variance for the constant arrival process and the Bernoulli process with arrival probability $P_{\text{arr}}s = 0.7$ is much smaller as compared to the variance for the Bernoulli process with arrival probability $P_{\text{arr}} = 0.1$. As the arrival probability decreases, the variance of the system energy with random arrivals decreases slowly with the number of users. This is due to the fact that the system energy converges to its mean value when approximately the same amount of data is scheduled in every time slot.

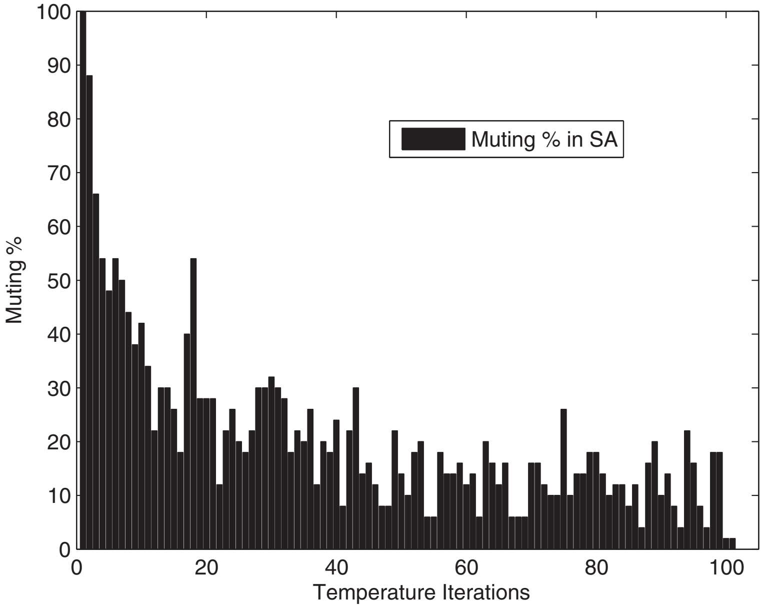

Figure 4 Muting update statistics for BA cooling schedule.

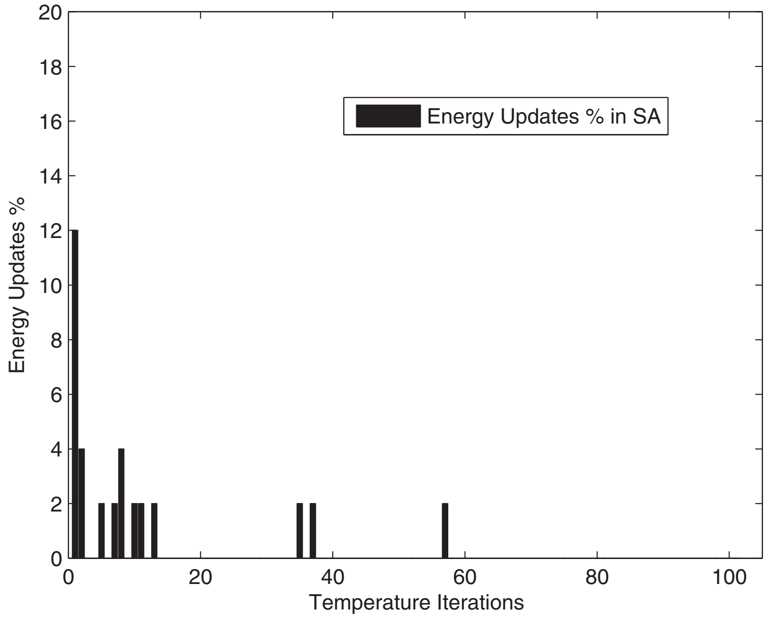

Figure 5 **Energy update statistics for BA cooling schedule.**

Obviously, a decrease in arrival probability results in a decrease in the amount of scheduled data and requires a larger number of users to compensate for this effect.

Figure 9 demonstrates the delay-energy trade-off for a single channel system when the arriving packets have non-identical deadlines. We evaluate the system performance at different spectral efficiencies. As the proportion of the packets with tight deadline constraints increases, the average system energy increases correspondingly as explained in Section 5. This effect is more pronounced at small spectral efficiencies.

We compare our scheduling scheme with the proportional fairness scheduler (PFS) proposed in [2]. PFS does not provide any deadline guarantees. In PFS, the multiuser diversity gain scales with the number of users per channel K/M present in the system while there is no deadline

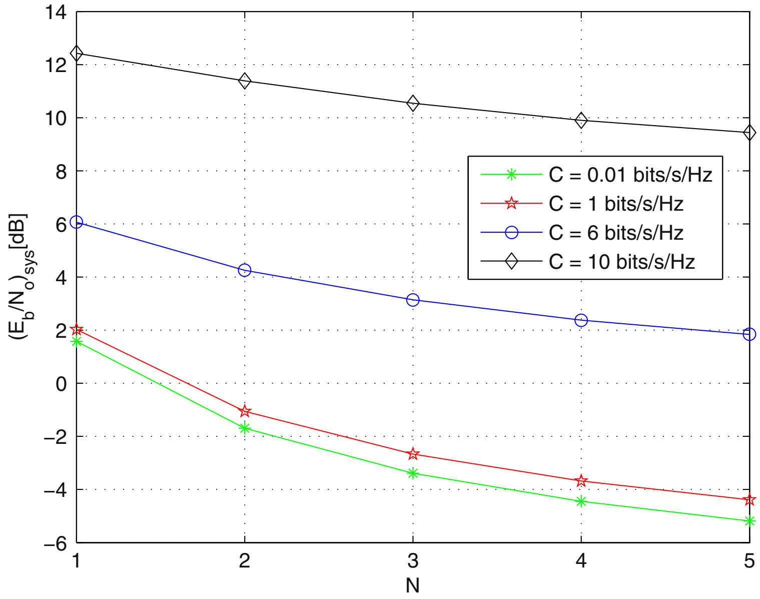

Figure 6 **The delay-energy behavior for the proposed scheduling scheme for a single channel system.**

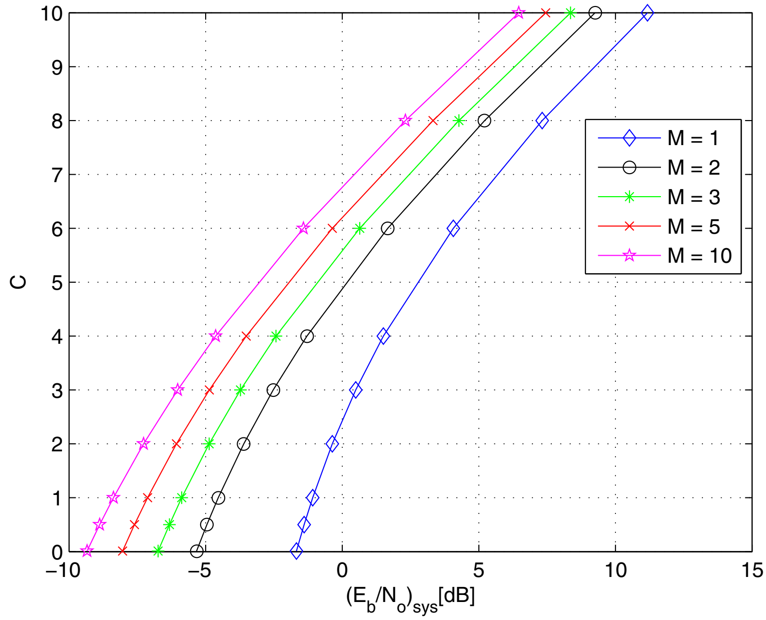

Figure 7 Impact of increase in number of channels on energy efficiency for the scheduling scheme.

delay constraint for the buffered packets. In our scheme, the multiuser diversity gain scales with the number of time slots available before reaching the deadline N while the number of users is asymptotically large. We refer to the parameters K/M and N as the degrees of freedom for the respective schemes. For a somehow fair comparison, we compare the two schemes for equal average delay in Figure 10. We use a Poisson arrival process for the evaluation of the delay behavior for both schemes. Figure 10 illustrates the average delay behavior of both schemes. The average delay of both PFS and our scheme scales linearly with increasing K/M and N, respectively. However, the average delay grows at a faster rate for PFS as compared to our scheme. Figure 11 compares the spectral efficiency of the two schemes. In general, PFS shows better results as compared to our scheme at small spectral efficiencies for the same degrees of freedom. However, our scheme outperforms PFS at large spectral efficiencies.

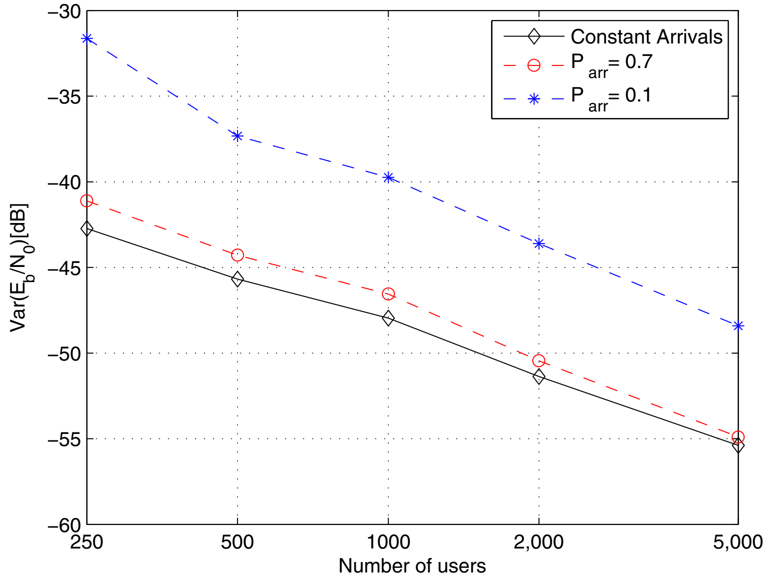

Figure 8 Effect of increase in number of users for the scheduling scheme with M = 10.

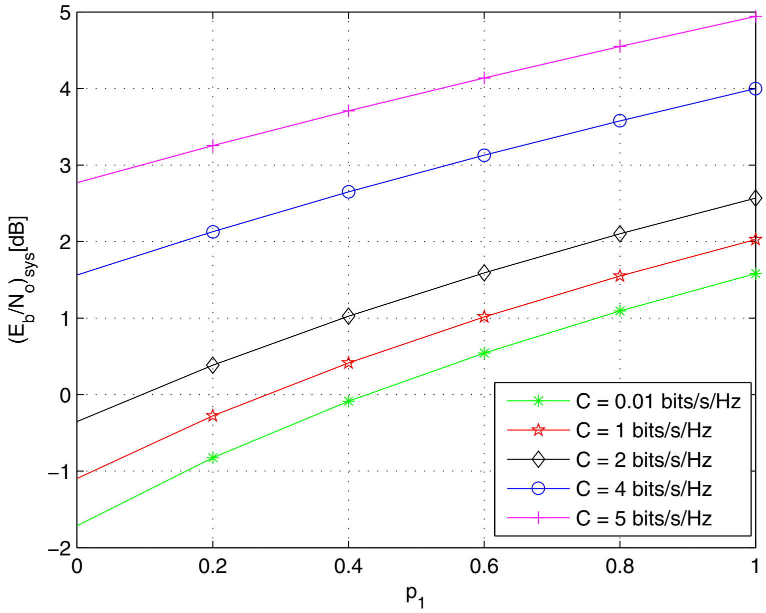

Figure 9 The arriving packets have deadline distances of 1 and 2 where $p_1 = 1 - p_2$. The results demonstrate the effect when probability of packets with deadline one increases.

For example, it beats PFS at $C = 2.3$ bits/s/Hz when the respective number of degrees of freedom equals 5. Furthermore, a comparison of the two schemes at the same average delay reveals further drawbacks of PFS. For example, we compare the two schemes for $M = 10$ and average delay of 2.5 time slots. PFS achieves this average delay at $K/M = 2$ ($K = 20$) while our scheme requires a deadline of $N = 5$ time slots as shown in Figure 10. A comparison of PFS with $K/M = 2$ to our scheme with $N = 5$ in Figure 11 shows that our scheme is able to beat PFS at even lower spectral efficiencies (1.5 bits/s/Hz as compared to 2.3 bits/s/Hz for equal degrees of freedom). At low spectral efficiency, PFS achieves better multiuser diversity gain as compared to delay-limited schemes and the cost of imposing delay constraint is high [1]. Thus, our scheme is more energy efficient than PFS at high spectral efficiencies

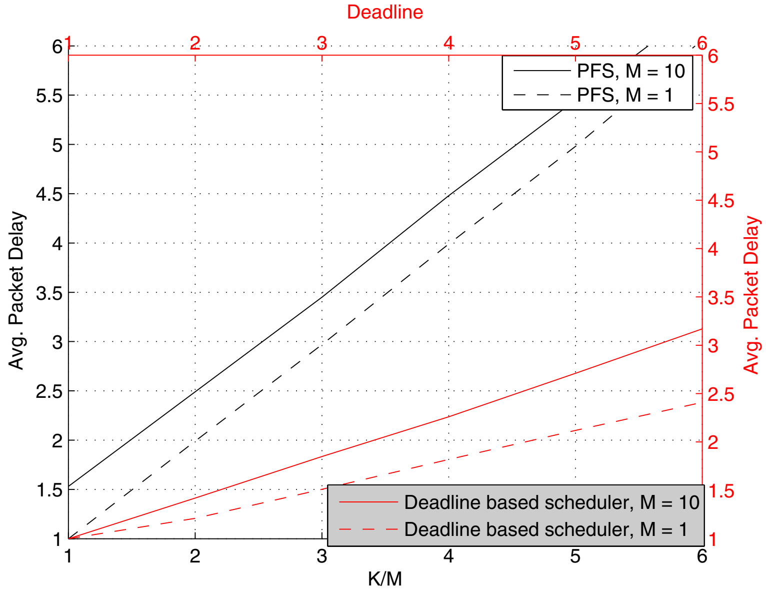

Figure 10 Comparison of PFS and our deadline-dependent scheduler in terms of average delay. Spectral efficiency *C* equals 0.5 bits/s/Hz.

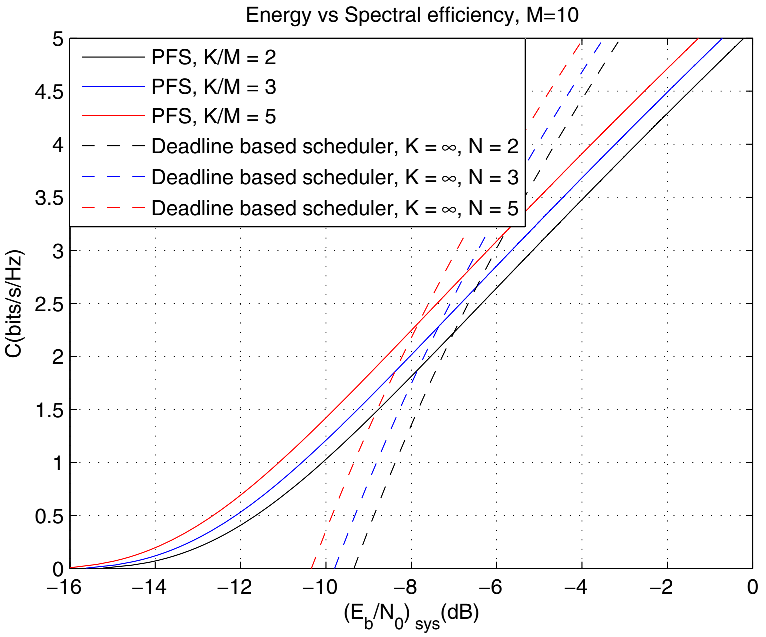

Figure 11 Comparison of PFS and our deadline dependent scheduler in terms of average system energy.

while it also provides better average delay performance at the same degrees of freedom.

9 Conclusion

We have proposed an energy efficient opportunistic multiuser scheduling scheme in the presence of a hard deadline delay constraint for the individual packets. The proposed scheme schedules the data depending on the instantaneous short-term fading and transmission deadline of the packets and exploits good channel conditions to make the system energy efficient. The many-user analysis and MDP modeling of the proposed scheme is the major contribution of this work. The many-user model helps to compute solution for the case of convex rate-power curve. Our system modeling ensures that the multiuser scheduling can be broken into a packet-based scheduling problem in the many-user limit. Though threshold optimization for the packet transmission is not a convex optimization problem, it can be solved within small margins of optimality with quite low complexity. We show that random arrivals can be modeled as constant arrivals with random size in the many-user limit and the scheduling decisions are independent of the deadline distribution of the arriving packets. The numerical results demonstrate that the many-user considerations are applicable for a reasonable network size of a few hundred users. The hard deadline can be used as a tuning parameter by the system designer to control the trade-off between the energy efficiency of the system and the maximum latency tolerated by the application.

Appendix

In this work, the channel model of [1] is used. Signal propagation is characterized by a distance-dependent path loss factor and a frequency-selective short-term fading that depend on the scattering environment around the user terminal. As described in Section 2, these two effects are taken into account by letting $g_k^m = s_k f_k^m$, where s_k denotes the path loss of user k and f_k^m is the short-term fading of user k in channel m.

As in [1], we assume that users are uniformly distributed in a geographical area but for a forbidden circular region of radius δ centered around the base station where $0 < \delta \le 1$ is a fixed system constant. Using this model, the cdf of path loss is given by

$$\mathrm{P}_s(x) = \begin{cases} 0 & x < 1 \\ 1 - \frac{x^{-2/\alpha} - \delta^2}{1-\delta^2} & 1 \le x < \delta^{-\alpha} \\ 1 & x \ge \delta^{-\alpha} \end{cases} . \qquad (28)$$

where the path loss at the cell border is normalized to one.

Frequency-selective short-term block fading is modeled by M parallel channel which are i.i.d. For a Rayleigh channel, the distribution of $f = \max\{f^1, \ldots, f^M\}$ is given by

$$\mathrm{P}_f(y) = (1 - \mathrm{e}^{-y})^M \qquad (29)$$

$\mathrm{P}_g(x)$ is defined as the cdf of the random variable $g = \max\{g^1, \ldots, g^M\} = s\max\{f^1, \ldots, f^M\}$. As path loss and

Rayleigh fading are statistically independent, the CDF of the channel gain is given by

$$\mathrm{P}_{g|d=1}(x) = \int \mathrm{P}_s(x/y)\,\mathrm{d}\mathrm{P}_{f|d=1}(y). \tag{30}$$

Using the path loss distribution in (28), (30) is computed as follows:

$$\mathrm{P}_{g|d=1}(x) = \int_0^{x\delta^\alpha} \mathrm{p}_{f|d=1}(y)\mathrm{d}y + \int_{x\delta^\alpha}^{x} \mathrm{P}_s(x/y)\,\mathrm{d}\mathrm{P}_{f|d=1}(y) \tag{31}$$

$$= \mathrm{P}_{f|d=1}(x\delta^\alpha) + \int_{x\delta^\alpha}^{x} \left(1 - \frac{(y/x)^{2/\alpha} - \delta^2}{1-\delta^2}\right) \mathrm{d}\mathrm{P}_{f|d=1}(y) \tag{32}$$

$$= \mathrm{P}_{f|d=1}(x) - \int_{x\delta^\alpha}^{x} \frac{(y/x)^{2/\alpha} - \delta^2}{1-\delta^2}\,\mathrm{d}\mathrm{P}_{f|d=1}(y). \tag{33}$$

Changing variables and integrating by parts yields,

$$\mathrm{P}_{g|d=1}(x) = \mathrm{P}_{f|d=1}(x) - \frac{\alpha}{2}\int_{x^{2/\alpha}\delta^2}^{x^{2/\alpha}} \frac{yx^{-2/\alpha} - \delta^2}{1-\delta^2}\, y^{\alpha/2-1}\, \mathrm{p}_{f|d=1}\left(y^{\alpha/2}\right)\mathrm{d}y \tag{34}$$

$$= \frac{x^{-2/\alpha}}{1-\delta^2}\int_{x^{2/\alpha}\delta^2}^{x^{2/\alpha}} \mathrm{P}_{f|d=1}\left(y^{\alpha/2}\right)\,\mathrm{d}y. \tag{35}$$

For $\alpha = 2$, (35) can be written in closed form. Using (25) and the Rayleigh fading model (29), (35) becomes

$$\mathrm{P}_{g|d=1}(x) = \frac{\left(1 + \sum_{i=0}^{N} \pi_i\left(1 - \mathrm{e}^{-\kappa_i}\right)^M\right)^{-1}}{x(1-\delta^2)} \times \sum_{i=0}^{N} \pi_i \int_{\max\{x\delta^2,\kappa_i\}}^{\max\{x,\kappa_i\}} \left(1 - \mathrm{e}^{-y}\right)^M + \left(1 - \mathrm{e}^{-\kappa_i}\right)^M\,\mathrm{d}y.$$

Following ([1], App. A), the closed form expression is given by

$$\mathrm{P}_{g|d=1}(x) = \frac{\left(1 + \sum_{i=0}^{N} \pi_i\left(1 - \mathrm{e}^{-\kappa_i}\right)^M\right)^{-1}}{x(1-\delta^2)} \times \sum_{i=0}^{N} \pi_i \left[\left(\max\{x,\kappa_i\} - \max\{x\delta^2,\kappa_i\}\right)\left(1 - \mathrm{e}^{-\kappa_i}\right)^M + \sum_{m=1}^{M} \frac{1}{m}\left(1 - \mathrm{e}^{-\max\{x,\kappa_i\}}\right)^m - \left(1 - \mathrm{e}^{-\max\{x\delta^2,\kappa_i\}}\right)^m\right].$$

Endnotes

[a] The dropping probability is defined as the probability that a packet cannot meet the deadline and dropped eventually after buffering for the time slots equal to deadline.

[b] The problem of error propagation in successive decoding can easily be overcome by means of iterative (soft) multiuser decoding [20].

[c] It should be noted that it may not be feasible to achieve deadline guarantee for every packet, e.g. due to shadowing or power limitation. The scheme can easily be extended to packet dropping scenario with non-zero dropping probability [25], but avoided here to focus on the main topic.

[d] It should be noted that computation of steady-state probabilities in a MDP requires solution of state equations with the condition $\sum_i \pi_i = 1$. Thanks to the tree structure of state diagram, we are able to compute the limiting probabilities in closed form *via* (17).

[e] The state in SA refers to the configuration of the system, i.e. the current transmission thresholds. It has no relation with the state of the Markov process given by the buffering time of the packet.

[f] Please note that optimization can also be performed for optimal $\vec{\alpha}$ as in Section 6.1 and computing optimal thresholds from $\vec{\alpha}^*$.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was supported by the Research Council of Norway (NFR) under the NORDITE/VERDIKT program (NFR contract no. 172177).

Author details

[1] Department of Computer Engineering, Qatar University, Doha, Qatar. [2] Institute of Digital Communications, Friedrich-Alexander University, Erlangen-Nuremberg, Germany. [3] Institute of Electronics and Telecommunications, Norwegian University of Science and Technology, Trondheim, Norway.

Received: 27 October 2013 Accepted: 6 April 2014
Published: 23 April 2014

References

1. G Caire, R Müller, R Knopp, Hard fairness versus proportional fairness in wireless communications: the single-cell case. IEEE Trans. Inform. Theory. **53**(4), 1366–1385 (2007)
2. P Viswanath, DNC Tse, R Laroia, Opportunistic beamforming using dumb antennas. IEEE Trans. Inform. Theory. **46**(6), 1277–1294 (2002)
3. RA Berry, RG Gallager, Communication over fading channels with delay constraints. IEEE Trans. Inform. Theory. **48**(5), 1135–1149 (2002)
4. D Wu, R Negi, Utilizing multiuser diversity for efficient support of quality of service over a fading channel. IEEE Trans. Veh. Tech. **54**(3), 1198–1206 (2005)
5. W Chan, MJ Neely, U Mitra, Energy efficient scheduling with individual packet delay constraints: offline and online results, in *IEEE Infocom* (IEEE, Piscataway, 2007)
6. MJ Neely, Optimal energy and delay tradeoffs for multiuser wireless downlinks. IEEE Trans. Inform. Theory. **53**(9), 3095–3113 (2007)
7. A Tarello, J Sun, M Zafar, E Modiano, Minimum energy transmission scheduling subject to deadline constraints. Wireless Networks. **14**(5), 633–645 (2008)
8. DP Bertsekas, *Dynamic Programming and Optimal Control, Vol. 1*, 3rd edn. (Athena Scientific, Nashua, 2007)

9. J Lee, N Jindal, Energy-efficient scheduling of delay constrained traffic over fading channels. IEEE Trans. Wireless Comm. **8**(4), 1866–1875 (2009)
10. P Viswanath, DN Tse, V Anantharam, Asymptotically optimal water-filling in vector multiple-access channels. IEEE Trans. Inform. Theory. **47**(1), 241–267 (2001)
11. M Benaïm, J-Y Le Boudec, A class of mean field interaction models for computer and communication systems. Perform Eval. **65**(11–12), 823–838 (2008)
12. MM Butt, EA Jorswieck, Maximizing system energy efficiency by exploiting multiuser diversity and loss tolerance of the applications. IEEE Trans. Wireless Comm. **12**(9), 4392–4401 (2013)
13. P Chaporkar, K Kansanen, RR Müller, On the delay-energy tradeoff in multiuser fading channels. EURASIP J. Wirel. Commun. Netw. **2009**, 1–14 (2009)
14. D Guo, S Verdu, Randomly spread CDMA: asymptotics via statistical physics. IEEE Trans. Inform. Theory. **51**(6), 1983–2010 (2005)
15. MM Butt, Energy-performance trade-offs in multiuser scheduling: large system analysis. IEEE Wireless Commun. Lett. **1**(3), 217–220 (2012)
16. N Jindal, S Vishwanath, A Goldsmith, On duality of Gaussian multiple access and broadcast channels. 768–783. **50**(5) (2004)
17. S Hanly, D Tse, Multi-access fading channels-part II: delay limited capacities. IEEE Trans. Inform. Theory. **44**(7), 2816–2831 (1998)
18. S ten Brink, G Kramer, A Ashikhmin, Design of low-density parity-check codes for modulation and detection. IEEE Trans. Comm. **52**(4), 670–678 (2004)
19. A Sanderovich, M Peleg, S Shamai, LDPC coded MIMO multiple access with iterative joint decoding. IEEE Trans. Inform. Theor. **51**(4), 1437–1450 (2005)
20. G Caire, R Müller, T Tanaka, Iterative multiuser joint decoding: optimal power allocation and low-complexity implementation. IEEE Trans. Inform. Theory. **50**(8), 1950–1973 (2004)
21. S Kirkpatrick, CD Gelatt, MP Vecchi, Optimization by simulated annealing. Science. **220**(4598), 671–680 (1983)
22. V Cerny, Thermodynamical approach to the travelling salesman problem: an efficient simulation algorithm. Optim. Theor. Appl. **45**(1), 41–52 (1985)
23. S Geman, D Geman, Stochastic relaxation, Gibbs distribution and the Bayesian restoration in images. IEEE Trans. Pattern Anal. Mach. Intell. **6**(6), 721–741 (1984)
24. H Szu, R Hartley, Fast simulated annealing. Phys. Lett. A. **122**(3–4), 157–162 (1987)
25. MM Butt, K Kansanen, RR Müller, Hard Deadline constrained multiuser scheduling for random arrivals, in *IEEE WCNC* (IEEE, Piscataway, 2011)

doi:10.1186/1687-1499-2014-65
Cite this article as: Butt *et al.*: Individual packet deadline delay constrained opportunistic scheduling for large multiuser systems. *EURASIP Journal on Wireless Communications and Networking* 2014 **2014**:65.