

# Energy-Performance Trade-offs in Multiuser Scheduling: Large System Analysis

M. Majid Butt, *Member, IEEE*

**Abstract**—This work addresses the problem of data scheduling over fading channels in a deadline (strict delay) constrained multiuser system where a certain amount of data loss can be tolerated. We propose a scheme which schedules the users based on their instantaneous channel conditions and backlog of the packets stored in the buffer. The scheme is analyzed in large user limit. Numerical results demonstrate the energy-performance trade-offs and quantify the gain in energy efficiency for the proposed scheme.

**Index Terms**—Multiuser diversity, energy-delay trade-off.

## I. INTRODUCTION

**D**ELAY guarantees and energy efficiency are the key features of any modern wireless communication system. It is important to design the system keeping into consideration the delay requirements of the applications. On the other hand, it is equally desirable to exploit the degrees of freedom available in terms of delay tolerance to achieve an energy efficient system. Traditionally, the focus of research has been to maximize capacity of the network. More recently, the focus has shifted to energy efficient communication due to high cost of energy in operating cellular networks and emergence of new applications like wireless sensor network.

This work discusses energy efficient scheduling for a fixed packet deadline constrained multiuser system. A deadline constrained packet can tolerate a fixed amount of delay before transmission. Multiuser diversity can effectively be used to minimize the system energy for such applications. A lot of researchers have worked on related resource allocation problems. The work in [1] deals with maximization of the information capacity by scheduling the users having relatively high instantaneous channel quality but this scheme does not provide any fairness guarantees. Proportional Fairness Scheduling (PFS) algorithm is proposed in [2] to provide fairness to the users. Energy-delay trade-off for deadline delay constrained systems is treated in [3], [4] where an individual packet deadline based scheduler is discussed that schedules a single user in each time slot. The proposed schemes do not quantify the packet loss due to poor channel conditions explicitly.

In this work, our goal is to develop an energy efficient multiuser scheduling scheme which incorporates the features of a practical scheduler and keeps the complexity low. We propose a scheduling scheme which guarantees hard deadline

delay for each user [5]. The loss tolerance of the application is an other important parameter that can be utilized to make the system more energy efficient. In practical systems, it is very difficult to provide hard deadline guarantees for all the packets. Some applications are loss tolerant and it is advantageous to drop certain proportion of the arriving packets to improve the system energy efficiency if the quality of the application remains acceptable. The gain in system energy efficiency depends on exploiting the loss and delay tolerance (before deadline) of the application in the scheduling process. We generalize the scheme in [5] for the loss tolerant systems and analyze it in large system limit. The large system analysis is essential to understand the independence of the users' scheduling decision process in a multiuser environment.

The rest of this paper is organized as follows. Sec. II introduces the system model used. Sec. III and IV discuss the proposed scheduling scheme and its asymptotic analysis, respectively. We evaluate the numerical results in Sec. V and conclude with the main contributions in Sec. VI.

## II. SYSTEM MODEL

We consider a multiple-access system with  $K$  (large) users randomly placed within a certain geographical area. Each user is provided a certain fraction of the total data rate available to the system. The required average rate  $R$  for each user is  $\frac{C}{K}$  where  $C$  denotes the spectral efficiency of the system.

We consider a time-slotted system. Each user  $k$  experiences a channel (power) gain  $g_k(t)$  in slot  $t$ . The channel gain  $g_k(t)$  is the product of path loss  $s_k$  and short-term fading  $f_k(t)$  i.e.  $g_k(t) = s_k f_k(t)$ . The path loss 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 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.

$E_k^R(t)$  and  $E_k(t)$  represent the received and the transmitted energy for each user  $k$  such that

$$E_k^R(t) = g_k(t)E_k(t). \quad (1)$$

Note that the distribution of  $g_k(t)$  differs from user to user. The channel state information is assumed to be known at both the transmitter and the receiver sides.

We allow multiple users to be scheduled in a time slot to avoid random dropping of the packets if multiple users reach the deadline simultaneously. The scheme follows the results for the asymptotic user case and therefore, there is no limit on the number of users scheduled simultaneously. Contrarily, there must be enough minimum scheduled users in every time slot so that the stochastic model of the scheduler

Manuscript received January 5, 2012. The associate editor coordinating the review of this letter and approving it for publication was R. M. Buehrer.

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

M. Majid Butt is with the Fraunhofer Heinrich Hertz Institute, Einsteinufer 37, 10555 Berlin, Germany (e-mail: majid.butt@ieee.org).

Digital Object Identifier 10.1109/WCL.2012.032312.120011

holds due to system level averaging. However, Monte Carlo Simulation based results in [5] verify that the proposed scheme is applicable for reasonably small number of scheduled users and remains practicable. The scheduled users are separated by superposition coding. Let  $\Delta$  be the set of users to be scheduled.  $\psi_k$  denotes the permutation of the scheduled user indices that sorts the channel gains in increasing order, i.e.  $g_{\psi_1} \leq \dots \leq g_{\psi_k} \leq \dots \leq g_{\psi_{|\Delta|}}$ . Then, the energy of the user  $\psi_k$  with rate  $R_{\psi_k}$ , as scheduled by the scheduler to guarantee an error free communication, is given by [6], [7]

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

where  $N_0$  denotes the noise power spectral density. Equation (2) represents the minimum total transmit energy assignment for the scheduled users.

### III. SCHEDULING ALGORITHM

The scheduling decision for a user  $k$  in a time slot  $t$  is based on the waiting time of the oldest packet in the buffer and instantaneous short-term fading  $f_k(t)$ . It is independent of the other users' scheduling decisions due to large number of users in the system. A set of users experiencing relatively large short-term fading is scheduled for transmission simultaneously. When a user is scheduled for transmission, she empties her buffer by scheduling all the packets buffered. Thus, the scheduling decision is simplified to a binary decision of transmitting the full buffer or waiting for the next time slot. If waiting time of the oldest buffered packet of a user is equal to the packet deadline  $n$ , the user is either scheduled for transmission or full buffer is dropped. This scheme is called Deadline Dependent Opportunistic Scheduling (DDOS).

As long as the waiting time of the oldest packet of the user is less than the deadline, the scheduler attempts to exploit the multiuser diversity gain and schedule the users in an energy efficient way. When the deadline is reached for the *oldest* packet in the buffer, the user is *forced* to transmit on the available channel or drop all the packets in the buffer. Note that emptying buffer property reduces the complexity of the scheduling decision which is an important factor in energy limited devices. Obviously, an individual packet deadline based scheduling algorithm is more energy efficient but it requires one threshold for every buffered packet. The additional complexity in online scheduling decision process can be significant for the network nodes with limited computation power. We propose a slightly more complex scheduler in Sec. IV-A that makes forced transmission of only the packet reaching the deadline. The numerical results in Sec. V show the improved energy efficiency at the cost of increased complexity.

The emptying buffer property requires that the duration of the time slot should be long enough that the transmission of all the buffered packets can be completed. In practical systems, this limits the maximum buffer size used. As we explain later, the buffer size models the deadline of the packets in our scheme. Thus, there is a practical limit on the maximum packet deadline that can be employed for this scheme depending on the channel model and the length of the time slot.

Note that we use fast fading instead of channel gain in the scheduling decision process to give every user equal opportunity of selection. Channel gain includes distance dependent path loss which inherently favours the users near the base station in the scheduling process.

#### A. Modeling of DDOS Scheduling

We use a Markov chain description to model and analyze the scheduler. We define some terms used in this work later.

**Definition 1 (Backlog State):** The backlog state (or simply state) in a Markov chain is defined as the waiting time of the oldest unscheduled packet in the buffer.

We consider random arrivals in each time slot for all the users and model them as constant arrivals with random content size. It has been proven in large system limit that such a representation results in equivalent state space descriptions for both constant and random arrivals due to system level averaging of data scheduled in each state [8]. This model makes our state space description independent of the arrival process. As a result of this model, backlog state represents the number of random size packets waiting to be scheduled in the buffer. Arrivals are queued in a finite buffer of  $n$  random size packets before transmission.

**Definition 2 (Opportunistic Threshold):** An opportunistic threshold  $\kappa_i$  is defined as the minimum short-term fading value allowing for the transition from state  $i$  to state 1.

It should be noted that computation of opportunistic thresholds for all the users depending on their respective backlog-states is usually not feasible in a multiuser environment because of a large state space. We consider an asymptotically large user system in this work. Therefore, the backlog-states of the users decouple and we can formulate the problem of energy efficient transmission in a multiuser system as an equivalent single user scheduling problem [9], [10].

In the Markov state description of DDOS, the deadline constraint is reflected by the maximum number of states  $n$ . State transition mechanism follows the following two properties.

- 1) If  $f \leq \kappa_i$ , no data is transmitted and forward state transition from a state  $i$  to the next higher state  $j$  occurs. As we model random arrivals in each time slot by a constant arrival of random size and assume identical deadline for all the arrived packets,  $j$  is always equal to  $i + 1$  when a user is not scheduled.
- 2) If  $f > \kappa_i$ , due to *emptying buffer* policy, the backward state transition occurs from state  $i$  to state 1 and the scheduler schedules  $i$  packets for transmission.

In a Markov process if a user is in state  $i$ , then the next state  $j$  is determined according to transition probabilities  $\alpha_{ij}$ . Let  $S_t$  be the state of the process at time  $t$ . Then, the state  $S_{t+1}$  depends only on state  $i$  and short-term fading  $f_k(t)$  at time  $t$ . The fading randomizes the state transitions. The resulting transition probabilities are given by

$$\alpha_{ij} = \Pr\{S_{t+1} = j | S_t = i\} = \begin{cases} \Pr(f > \kappa_i) & j = 1 \\ \Pr(f \leq \kappa_i) & j = i + 1 \\ 0 & \text{else} \end{cases} \quad (3)$$

$\alpha_{ij} = 0$  implies impossible state transition following the properties of the DDOS scheduler.  $\alpha_{n1}$  is partitioned into

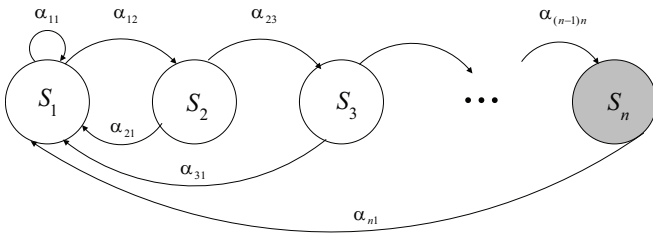


Fig. 1. State diagram for the transition states of a DDOS scheduler.

$\tilde{\alpha}_{n1}$  and  $\hat{\alpha}_{n1}$  where  $\tilde{\alpha}_{n1}$  and  $\hat{\alpha}_{n1}$  denote the probabilities of scheduling and dropping the buffer in state  $n$ , respectively.  $\hat{\alpha}_{n1}$  is computed using predefined application dependent dropping probability  $\theta$  such that

$$\hat{\alpha}_{n1} = \Pr(f \leq \kappa_n) = \frac{\theta}{\pi_n} \quad (4)$$

where  $\pi_n$  is the limiting probability of state  $n$  in a Markov chain. Equation (4) states that a fraction  $\theta$  of all the packets entering the system is dropped in state  $n$ . Then,  $\tilde{\alpha}_{n1}$  is computed as

$$\tilde{\alpha}_{n1} = \Pr(f > \kappa_n) = 1 - \frac{\theta}{\pi_n} \quad (5)$$

The state transition diagram is shown in Fig. 1 and the corresponding transition probability matrix is given by

$$\mathbf{P} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & 0 & \cdots & 0 \\ \alpha_{21} & 0 & \alpha_{23} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{\alpha}_{n1} + \hat{\alpha}_{n1} & 0 & 0 & \cdots & 0 \end{pmatrix}. \quad (6)$$

where  $\tilde{\alpha}_{n1} + \hat{\alpha}_{n1}$  equals one due to the *forced* packet transmission or dropping in state  $n$ .

We assume that the users exhibit independent fading processes. Furthermore, the proposed scheduler is independent of the other users' fading. The law of large numbers drives the proportion of users in state  $i$  at time  $t$  to be identical to  $\Pr(S_t = i)$  in the large user limit.

If a user is scheduled in a small queue state, it utilizes a good channel. If the channel is not good, the user may decide to wait for the next slot. However, decision to wait for the next time slot brings the deadline for the user closer and the user runs the risk of scheduling full buffer in a potentially bad channel in state  $n$ . The task of the optimizer is to find the optimal threshold values such that the transmitted energy in the system is minimized by smart scheduling.

#### IV. LARGE SYSTEM ANALYSIS OF DDOS

To analyze DDOS in large system limit, we apply the results of [7]. For analysis purpose, we model a user that sends  $L$  packets at a time as  $L$  virtual users with identical fading that send single packets. Thus, we denote the scheduled packets by single virtual users (SVU). The average energy consumption of the system per transmitted information bit at the large system limit  $K \rightarrow \infty$  is then given by [7]

$$\left(\frac{E_b}{N_0}\right)_{\text{sys}} = \log(2) \int_0^\infty \frac{2^C P_{g,\text{SVU}}(x)}{x} dP_{g,\text{SVU}}(x) \quad (7)$$

where  $P_{g,\text{SVU}}(\cdot)$  denotes the cumulative distribution function (cdf) of the channel gains of the SVUs. It is composed of the short-term fading and the long-term fading of the SVUs.

In the following, we evaluate probability distribution function (pdf) of the channel gain  $p_{g,\text{SVU}}(x)$  of the scheduled virtual users. The scheduling decisions affect the short-term fading distribution only. The resulting pdf of the short-term fading of the SVUs is given by

$$P_{f,\text{SVU}}(y) = \begin{cases} \sum_{i=1}^n c_i \pi_i i p_f(y) & y > \kappa_i \\ 0 & \text{else} \end{cases} \quad (8)$$

where  $p_f(y)$  and  $c_i$  denote the short-term fading distribution and a constant to normalize the pdf.

Using (8), the cdf of the SVUs is given by

$$P_{f,\text{SVU}}(y) = \sum_{i=1}^n c_i \pi_i i (P_f(y) - P_f(\kappa_i)). \quad (9)$$

$P_{g,\text{SVU}}(y)$  is calculated using (9) and path loss distribution.

#### A. Improvement in Energy Efficiency

The emptying buffer policy simplifies the scheduling decision process. However, when the oldest buffered packet reaches the deadline, the forced scheduling or dropping of the full buffer results in performance loss. To mitigate this loss, a simple modification works as follows.

In deadline state, the user empties the buffer only if fading is better than the threshold  $\kappa_n$  as before but this threshold does not depend on dropping probability  $\theta$ . If  $f < \kappa_n$ , we decide to do forced transmission or dropping of *only* the packet reaching the deadline. The other packets remain buffered and the scheduler returns back to state  $n$ . To decide transmission of the oldest buffered packet, we require an additional dropping probability dependent threshold  $\kappa_M$ . It is computed from  $\hat{\alpha}_{nn}$  and  $\tilde{\alpha}_{nn}$  which denote the transition probabilities for dropping and scheduling a single packet, respectively. This scheduler is called Modified Deadline Dependent Scheduler (MDDOS).

#### B. Optimization of Thresholds

In this section, we outline the procedure for the optimization of thresholds. The vector of thresholds is computed offline using the statistics of short-term fading. For a deadline of  $n$  time slots, we require  $n$  and  $n+1$  thresholds for DDOS and MDDOS, respectively. Note that the last deadline is a function of the other thresholds and dropping probability  $\theta$  as explained in Section III. The vector of thresholds  $\vec{\kappa}$  is optimized using (6) and Simulated Annealing (SA) algorithm with system energy as an objective function. The transmit energy is not a convex function of the threshold vector. Therefore, we require some heuristic approach to optimize  $\vec{\kappa}$ . In [5], a heuristic recursive algorithm has been proposed to optimize the thresholds. However, we discovered that this algorithm is not accurate for the case when  $\theta > 0$  and SA provides better solution than the recursive algorithm in [5]. The SA algorithm is believed to provide near optimal solutions in many non-convex problems e.g. traveling salesman problem. We skip the details of SA due to space limitation. The interested reader is referred to [8], [11] for the detailed discussion on SA.

TABLE I  
THRESHOLD COMPUTATION FOR DDOS AT  $C = 0.5$  BITS/S/Hz

$n$	$\theta = 0$					$\theta = 0.1$				
	$\kappa_1$	$\kappa_2$	$\kappa_3$	$\kappa_4$	$E_b/N_0$ [dB]	$\kappa_1$	$\kappa_2$	$\kappa_3$	$\kappa_4$	$E_b/N_0$ [dB]
2	0.16	0	-	-	-0.76	0.57	0.40	-	-	-4.43
3	0.36	0.18	0	-	-2.24	1.0	0.90	0.76	-	-6.01
4	0.56	0.40	0.19	0	-3.23	1.48	1.33	1.23	1.15	-7.09

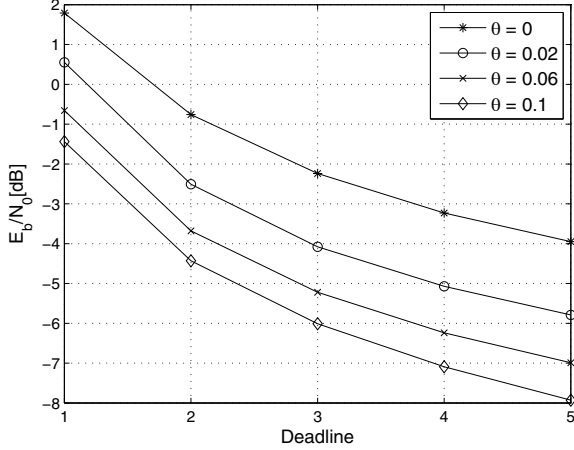


Fig. 2. Energy-delay trade-off for a multiuser system at  $C = 0.5$  bits/s/Hz.

## V. NUMERICAL RESULTS

We consider a multiuser multi-access system. The 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 exponent  $\alpha$  equals 2 and path loss distribution follows the model described in [7]. All the users experience fast fading with exponential distribution with mean one.

Table I shows the optimized thresholds for the DDOS scheduler for different deadlines and dropping probabilities using SA algorithm. Note that  $\kappa_n$  corresponds to the dropping probability  $\theta$ . For the same  $n$ , an increase in the value of  $\theta$  results in increase in the values of threshold and decrease in probability of costly forced transmission. Thus, the scheduler has more freedom to wait for a better channel in the future.

Fig. 2 exhibits the performance-delay trade-off for different values of deadline delay  $n$  and dropping probability  $\theta$ . Both parameters are the forms of diversity provided by the characteristic of the application and work as degrees of freedom. An increase in the value of deadline delay constraint or/and dropping probability makes the system more energy efficient. The  $E_b/N_0$  values are quite low because the users exploit multiuser diversity inherent in the system. As the results are based on the large user limit, it implies that there will always be high number of scheduled users with relatively large channel gains that contribute to reduce the system energy.

Fig. 3 shows the comparison of DDOS and MDDOS schemes. We observe an improvement in system energy for MDDOS as compared to DDOS for all deadlines and spectral efficiencies. However, MDDOS requires an additional threshold and complexity of implementation of the scheduler needs to be considered to determine the over all energy efficiency of the scheduler.

## VI. CONCLUSION

We propose an opportunistic multiuser scheduling scheme to provide a deadline guarantee to the users at low system

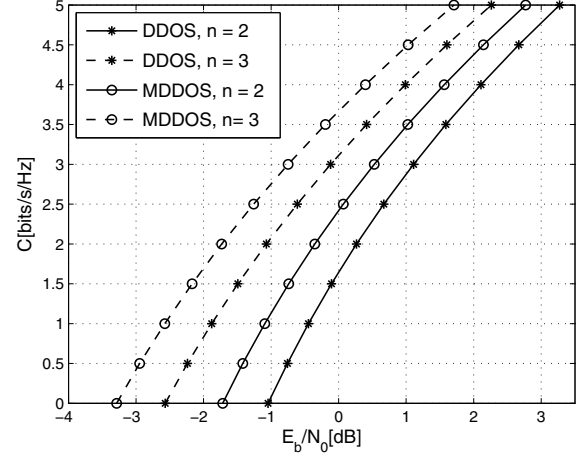


Fig. 3. Comparison of DDOS and MDDOS for a system with  $\theta = 0$ .

energy and small computational complexity. The main contribution of this work is the large system analysis of the proposed scheduling scheme for the hard deadline constrained multiuser systems. We evaluate the proposed scheme for the loss tolerant applications explicitly and conclude that delay and loss tolerance are important degrees of freedom to design an energy efficient radio resource allocation scheme.

## REFERENCES

- [1] R. Knopp and P. Humblet, "Information capacity and power control in single cell multiuser communications," in *1995 IEEE Int. Computer Conf.*
- [2] P. Viswanath, D. N. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 1277–1294, June 2002.
- [3] A. Tarello, J. Sun, M. Zafar, and E. Modiano, "Minimum energy transmission scheduling subject to deadline constraints," *Wireless Networks*, vol. 14, no. 5, pp. 633–645, Oct. 2008.
- [4] J. Lee and N. Jindal, "Energy-efficient scheduling of delay constrained traffic over fading channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1866–1875, Apr. 2009.
- [5] M. M. Butt, K. Kansanen, and R. R. Müller, "Provision of maximum delay guarantee at low energy in a multiuser system," in *2008 International Symposium on Spread Spectrum Techniques and Applications*.
- [6] D. Tse and S. Hanly, "Multi-access fading channels—part I: polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2796–2815, Nov. 1998.
- [7] G. Caire, R. Müller, and R. Knopp, "Hard fairness versus proportional fairness in wireless communications: the single-cell case," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1366–1385, Apr. 2007.
- [8] M. M. Butt, K. Kansanen, and R. R. Müller, "Hard deadline constrained multiuser scheduling for random arrivals," in *2011 WCNC*.
- [9] M. Benaïm and J.-Y. Le Boudec, "A class of mean field interaction models for computer and communication systems," *Performance Evaluation*, vol. 65, no. 11–12, pp. 823–838, 2008.
- [10] P. Viswanath, D. N. Tse, and V. Anantharam, "Asymptotically optimal water-filling in vector multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 1, pp. 241–267, Jan. 2001.
- [11] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, May 1983.