DTU Library

Priority-based Resource Allocation for RT and NRT Traffics in OFDMA Systems

Wang, Hua

[Link back to DTU Orbit](#)

Priority-based Resource Allocation for RT and NRT Traffics in OFDMA Systems

Hua Wang

Department of Communications, Optics & Materials
Technical University of Denmark, Lyngby, Denmark
Email: huw@com.dtu.dk

Abstract— In this paper, we address the problem of adaptive radio resource allocation and QoS provisioning for real-time (RT) and non-real-time (NRT) services in OFDMA systems. The proposed algorithm tightly couples the subchannel allocation and packet scheduling together through an integrated cross-layer approach in which each packet is assigned a priority value by taking both the instantaneous channel condition and the QoS requirements into account. An efficient suboptimal algorithm with low computational complexity is proposed to solve the linear optimization problem. Simulation results show that the proposed algorithm can greatly improve the system performance in terms of high spectral efficiency and low outage probability compared to the traditional TDMA algorithm, thus is very suitable for the downlink of current OFDMA systems.

I. INTRODUCTION

Orthogonal Frequency Division Multiple Access (OFDMA) is an attractive multiple access scheme for future wireless and mobile communication systems due to its excellent performance over multipath fading channels. In OFDMA systems, the carrier bandwidth is divided into multiple subcarriers, and each user is dynamically assigned to a subset of subcarriers. This capability enables the network to perform a flexible resource management, such as subcarrier allocation, adaptive modulation, and power distribution, with the goal of enhancing the resource utilization efficiency and providing diverse quality of service (QoS) guarantees under different traffic loads and time-varying channel conditions.

Recently, resource management in OFDMA systems has attracted enormous research interests. [1]-[3] addressed the problem of minimizing the total transmitted power at a given bit rate per terminal. Subcarrier and bit allocation were done dynamically through the use of nonlinear optimization with integer variables. The alternative objective of maximizing the overall system throughput with upper bounded transmitted power has been considered in [4]-[6]. These schemes allow each user to specify its individual QoS requirement, defined in terms of a fixed data rate per frame.

However, in practical communication systems, it is neither sufficient nor efficient to represent the QoS requirement by a fixed data rate per frame. The resource allocation problem for supporting both real-time (RT) and non-real-time (NRT) traffic in multimedia systems becomes much more complicated when diverse QoS requirements have to be considered. The transmission of RT packets can be delayed as long as the delay constraint is not violated, and the transmission of NRT packets

can be more elastic. Therefore, for efficient utilization of radio resources, the bandwidth allocation should be controlled dynamically and adaptively. This is pursued in [7]-[8].

In this paper, we propose a new resource allocation algorithm for supporting both RT and NRT traffics in OFDMA systems. Unlike the conventional approaches, which decompose the resource allocation into two steps: packet scheduling and subcarrier-and-power allocation [3] & [8], the proposed algorithm tightly couples these two processes together through an integrated cross-layer approach to take advantage of the inter-dependencies between the PHY and MAC layers. The idea is that we assign a priority value based on a priority function to each packet if it is scheduled for transmission on a specific subcarrier. Then we can formulate the resource allocation problem into an optimization problem with the objective of maximizing the total achievable priority values. Simulation results show that the proposed algorithms can yield high spectral efficiency with satisfying QoS requirements.

The rest of this paper is organized as follows. The system model and resource allocation problem is formulated in Section II. Section III presents a suboptimal computation-complexity reduced algorithm in details. Simulation environments and results are outlined and discussed in Section IV. Finally, a conclusion is drawn in Section V.

II. RESOURCE ALLOCATION MODEL

We consider the downlink of an OFDMA system with N subchannels and K users. The time axis is divided into frames. A frame is further divided into S time slots, each of which may contain one or several OFDM symbols. The duration of a frame is set to be 1 ms, thus we can assume that the channel quality remains constant within a frame, but may vary from frame to frame. Two separate queues, a RT queue and a NRT queue, are maintained for each user at the base station (BS). Each queue has a finite length of L packets. On arriving at the BS, the packets are buffered in the corresponding queue and are served by the BS scheduler. We assume that packets are of fixed length by fragmentation in the MAC layer and each packet contains d information bits.

The resource allocation at the BS involves the allocation of subchannels, time slots and transmission powers. The basic idea behind the proposed algorithm is that at each scheduling interval, if a packet was scheduled for transmission, it is assigned a priority value based on the instantaneous channel

condition (PHY layer issue), as well as the QoS requirement (MAC layer issue). The objective of our resource allocation algorithm is to maximize the total achievable priority values, thus maximize the system throughput with satisfying QoS requirements for both RT and NRT traffics.

We apply the extended Exponential rule (EXP) [9] as our priority function. It was proposed to provide QoS guarantees over a shared wireless link in terms of the average packet delay for RT traffic, and a minimum throughput for NRT traffic.

For RT traffic, if the ith packet from the jth queue of user k is scheduled for transmission on subchannel n, its priority value is calculated as:

$$P(k,j,i,n) = a_k \cdot \frac{\mu_{k,n}(t)}{\overline{\mu}_k(t)} \cdot \exp\left(\frac{a_k W_{k,j,i}(t) - \overline{aW}}{1 + \sqrt{\overline{aW}}}\right) \quad (1)$$
$$j \in \text{RT}$$

where $\overline{aW} = \frac{1}{K}\sum_k a_k W_{k,j,1}(t)$, and $a_k = -\log \delta_k / T_{k,\max}$. $W_{k,j,i}(t)$ is the ith packet delay from the jth queue of user k at time t, $T_{k,\max}$ is the maximum allowable delay of user k, δ_k is the maximum outage probability of user k, $\mu_{k,n}(t)$ is the instantaneous channel rate if subchannel n is assigned to user k at time t, and $\overline{\mu}_k(t)$ is the average channel rate of user k over a sliding window size t_c, calculated as:

$$\overline{\mu}_k(t) = (1 - \frac{1}{t_c})\overline{\mu}_k(t-1) + \frac{1}{t_c}\mu_k(t) \quad (2)$$

where $\mu_k(t) = \sum_{n=1}^{N} c_{k,n} \cdot \mu_{k,n}(t)$ is the total channel rate of user k at time t. If subchannel n is assigned to user k, $c_{k,n} = 1$, otherwise $c_{k,n} = 0$.

For NRT traffic, the EXP rule is used in conjunction with a token bucket control to guarantee a minimum throughput [9]. We associate each NRT queue with a virtual token bucket. Tokens in each bucket arrive at a constant rate $r_{k,\text{req}}$, which is the required minimum throughput to user k. After a packet is scheduled for service, the number of tokens in the corresponding token queue is reduced by the actual amount of data transmitted. The calculation of the priority value for a NRT packet is similar to Exp. (1), with the exception that $W_{k,j,i}(t)$ in NRT is the virtual waiting time, defined as:

$$W_{k,j,i}(t) = \frac{\max\{0, V_{k,j}(t) - (i-1)\cdot d\}}{r_{k,\text{req}}} \quad j \in \text{NRT} \quad (3)$$

where $V_{k,j}(t)$ is the number of tokens associated with the jth NRT queue of user k at time t, $r_{k,\text{req}}$ is the minimum required throughput of user k, and d is the packet length.

Let us define $u(k,j,i,n)$ be the subchannel allocation indicator. That is, $u(k,j,i,n) = 1$ means that the ith packet from the jth queue of user k is allocated to subchannel n for packet transmission, and $u(k,j,i,n) = 0$ otherwise. Also let's define $p(k,n)$ and $g(k,n)$ be the transmission power allocated to user k on subchannel n, and the channel gain perceived by user k on subchannel n, respectively. Then the achievable channel rate when user k is assigned to subchannel n can be written as [4]:

$$\mu_{k,n}(t) = \Delta B \cdot \log_2\left(1 + \frac{-1.5 \cdot \text{SINR}_{k,n}}{\ln(5 \cdot \text{BER}_k)}\right) \quad (4)$$

where $\text{SINR}_{k,n}$ is the kth user's received signal-energy-to-noise-density on subchannel n, which is defined as:

$$\text{SINR}_{k,n} = \frac{g_{k,n} \cdot p_{k,n}}{N_0 \cdot \Delta B} \quad (5)$$

where ΔB is the subchannel bandwidth. BER_k is the Bit Error Rate (BER) requirement of user k, and N_0 is the noise power spectral density.

Let's define $m(k,j,i,n)$ be the number of time slots needed on subchannel n if the ith packet from the jth queue of user k is scheduled for transmission on subchannel n, expressed as:

$$m(k,j,i,n) = \frac{d}{\mu_{k,n}(t) \cdot \Delta t} \quad (6)$$

where Δt is the length of one time slot.

Then, the objective of the adaptive resource allocation problem can be mathematically formulated as follows:

$$\arg \max_{u(k,j,i,n)} \sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{i=1}^{L}\sum_{n=1}^{N} u(k,j,i,n) \cdot P(k,j,i,n) \quad (7)$$

subject to:

$$\sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{i=1}^{L} u(k,j,i,n) \cdot m(k,j,i,n) \leq S \quad \forall n \quad (8)$$

$$\sum_{n=1}^{N} u(k,j,i,n) \leq 1 \quad \forall k,j,i \quad (9)$$

$$u(k,j,i,n) \in \{0,1\} \quad \forall k,j,i,n \quad (10)$$

We assume that the total transmission power P_0 is equally distributed among all subchannels, that is, $p(k,n) = P_0/N$.

The above optimization problem can be solved by determining the values of integer variables $u(k,j,i,n)$ through standard linear integer programming. The solution to the problem provides an optimal resource allocation. However, the computation complexity of the optimal solution is too high to be applied in practical systems. To reduce the computation complexity, we propose a suboptimal algorithm in the next section.

III. PROPOSED SUBOPTIMAL SCHEME

In the suboptimal algorithm, we allocate radio resources on a packet by packet basis. The general idea is that, at the scheduling interval, the packet with the highest priority value from all queues is scheduled for transmission, and this procedure iterates until either there is no radio resource left or there is no packet remaining unscheduled in the queue. A detailed description of the resource allocation algorithm is listed in pseudocode 1, where Ω_s^k is the set of subchannels available for data transmission for user k, Ω_q^k is the set of queues having pending traffic of user k, s_n is the number of residual time slots on subchannel n, $i_{k,j}$ is a pointer to the next packet to be scheduled in the jth queue of user k, and L is the maximum queue length.

If user k has pending traffic, the proposed algorithm first pre-allocates the best subchannel n in terms of the channel

Algorithm 1 Suboptimal Resource Allocation Algorithm

1: Set $s_n \leftarrow S$ for $\forall n$ {initialize s_n}
2: Set $i_{k,j} \leftarrow 1$ for $\forall k, j$ {initialize $i_{k,j}$}
3: Set $\Omega_s^k \leftarrow \{1, \cdots, N\}$ for $\forall k$ {initialize Ω_s^k}
4: Set Ω_q^k to be set of queues that are not empty of user k
5: **while** $\exists x, \Omega_s^x \neq \phi$ **do**
6:   **for** $k = 1$ to K **do**
7:     **while** $\Omega_s^k \neq \phi$ **do**
8:       **if** $\Omega_q^k \neq \phi$ **then**
9:         Select $n \leftarrow \arg\max_{n \in \Omega_s^k} \mu_{k,n}(t)$ {assign the best channel from user's available channel set}
10:         **if** $s_n \geq \frac{d}{\mu_{k,n}(t)}$ **then**
11:           **for** every $j \in \Omega_q^k$ **do**
12:             Calculate $P(k, j, i_{k,j}, n)$ in Exp. (1)
13:           **end for**
14:           BREAK
15:         **else**
16:           $\Omega_s^k \leftarrow \Omega_s^k - \{n\}$ {remove n from the channel set if there is not enough capacity left}
17:           CONTINUE
18:         **end if**
19:       **else**
20:         $\Omega_s^k \leftarrow \phi$ {set Ω_s^k to be null if Ω_q^k is null}
21:       **end if**
22:     **end while**
23:   **end for**
24:   Schedule packet $(k^*, j^*, i_{k,j}^*)$ on subchannel n^*, where $(k^*, j^*, i_{k,j}^*, n^*) \leftarrow \arg\max P(k, j, i_{k,j}, n)$
25:   $s_{n^*} \leftarrow s_{n^*} - \frac{d}{\mu_{k^*, n^*}}$
26:   **if** $i_{k^*, j^*} = L$ **then**
27:     $\Omega_q^{k^*} \leftarrow \Omega_q^{k^*} - \{j^*\}$ {remove queue j^* from the queue set if all packets in the queue have been scheduled}
28:   **else**
29:     $i_{k^*, j^*} \leftarrow i_{k^*, j^*} + 1$
30:   **end if**
31: **end while**

quality to user k from his available subchannel set. If there is not enough capacity left on the best subchannel n to accommodate one packet from user k's queue, that is, if $s_n < d/\mu_{k,n}(t)$, subchannel n will be removed from user k's available subchannel set and the second best subchannel n' will be selected. This procedure continues until a best possible subchannel is pre-allocated to user k. Otherwise, user k is removed from the list. After the subchannel pre-allocation process for all users is done, the algorithm calculates the priority value of the head-of-line (HOL) packet in each non-empty queue, and schedule the packet with the highest priority value $(k^*, j^*, i_{k,j}^*, n^*) = \arg\max P(k, j, i_{k,j}, n)$ for transmission on subchannel n^*. The scheduled packet is removed from the corresponding queue and the consumed radio resources in terms of time slots are subtracted on subchannel n^*. Then it starts from the beginning and continues until either there is no radio resource left or no packets pending in the queue.

IV. SIMULATION RESULTS AND DISCUSSIONS

We evaluate the performance of the proposed scheduling algorithm by a system-level simulation in OPNET. We consider an OFDMA downlink system with cell radius of 2 km, where subscriber stations (SSs) are randomly placed in the cell with uniform distribution. The total bandwidth is set to be 5 MHz, which is divided into $N = 10$ subchannels. Adaptive modulation and coding scheme is applied to transmit data on each subchannel such that the highest possible rate is chosen. We consider pass loss, shadowing, and frequency selective Rayleigh fading as channel models. We assume that perfect channel information is known at BS. Three traffic types are considered. Each user generates one or several traffic types independently. VoIP and videoconference are served in RT class while Web is served in NRT class. A summary of traffic parameters for different traffic types are listed in Table I. The delay constraint for RT class is 50 ms and the minimum throughput constraint for NRT class is 100 Kbits/sec. For QoS evaluation, we measure the outage probabilities for RT and NRT classes respectively. Fragmentation is enabled where the maximum MAC PDU size is set to be 56 bytes. For comparisons, we include the simulation results of TDMA/EXP rule as it is considered to be one of the best TDMA based scheduling algorithms, of which each user transmits in the assigned time slots over the entire allocated bandwidth.

Type	Characteristics	Distribution	Parameters
VoIP	ON period	Exponential	Mean = 1.34 sec
VoIP	OFF period	Exponential	Mean = 1.67 sec
VoIP	Packet size	Constant	66 bytes
VoIP	Inter-arrival time between packets	Constant	20 ms
Video	Packet size	Log-normal	Mean = 4.9 bytes Std. dev. = 0.75 bytes
Video	Inter-arrival time between packets	Normal	Mean = 33 ms Std. dev. = 10 ms
Web	Reading time between sessions	Exponential	Mean = 5 sec
Web	Number of packets within a session	Geometric	Mean = 25 packets
Web	Inter-arrival time between packets	Geometric	Mean = 0.0277 sec
Web	Packet size	Truncated Pareto	$k = 81.5$ bytes $\alpha = 1.1$ $m = 2$ M bytes

TABLE I

A SUMMARY OF TRAFFIC PARAMETERS

Fig. 1 depicts the average packet delay in RT service of three schemes, which shows that the TDMA/EXP scheme has a much longer packet delay than the proposed scheme, which results in a higher delay outage probability as shown in Fig. 2. The throughput performance of NRT service is depicted in Fig. 3. The proposed scheme achieves a much higher throughput than the TDMA/EXP scheme, thus has a lower throughput outage probability. Fig. 4 depicts the normalized spectral efficiency, which is defined as the ratio between the achieved modulation and the highest modulation, under three schemes. It can be seen that the optimal scheme and
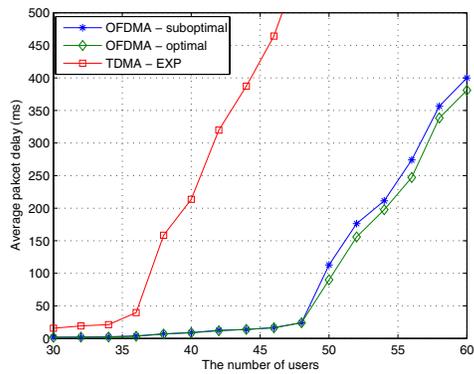
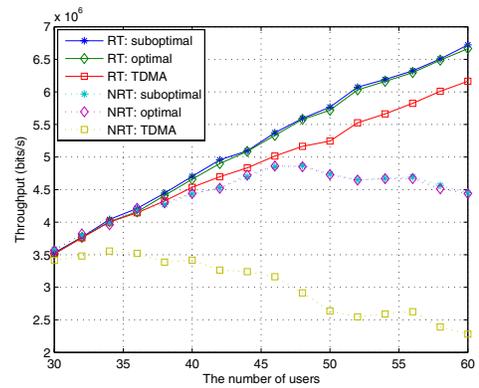Fig. 1. Average packet delay in RT

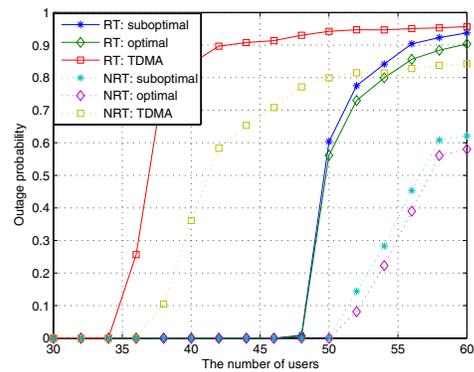Fig. 3. Average throughput in RT and NRT

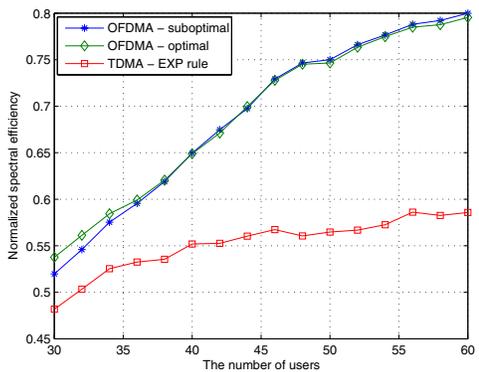
Fig. 2. Outage probability in RT and NRT

Fig. 4. Normalized spectral efficiency in RT and NRT

the suboptimal scheme achieve similar spectral performance, while it is much lower in the TDMA/EXP scheme. This is because our proposed scheme not only exploits multiuser diversity in the time domain, but also in the frequency domain. Numerical results from the simulations show that the proposed suboptimal scheme performs close to the optimal scheme, but with considerably low computation complexity.

V. CONCLUSIONS

This paper addresses the problem of QoS scheduling and resource allocation for real-time (RT) and non-real-time (NRT) services in multiuser OFDMA systems. The proposed algorithm assigns a priority value to each packet based on the extended EXP rule which tightly couples the PHY layer and MAC layer issues together. To reduce the computational complexity of a linear optimization problem, a suboptimal algorithm is proposed. Through system-level simulation, the performance of the proposed algorithms is evaluated in terms of packet delay, throughput, outage probability and spectral efficiency in both RT and NRT services. It is shown that the performance of the suboptimal scheme is slightly different from the optimal scheme, and both schemes achieve higher spectral efficiency, throughput, and lower packet delay, outage probability compared with the conventional TDMA scheme.

REFERENCES

[1] Ying J. Z., and Khaled B.: *Energy-Efficient MAC-PHY Resource Management with Guaranteed QoS in Wireless OFDM Networks*, ICC 2005, Vol.5, pp. 3127–3131, May. 2005.
[2] Ying J. Z., and Khaled B.: *Adaptive Resource Allocation and Scheduling for Multiuser Packet-based OFDM Networks*, ICC 2004, Vol.5, pp. 2949–2953, June. 2004.
[3] Alfedo T., Marco M., Andrea V. and Andrea B.: *A modular cross-layer scheduling and resource allocation architecture for OFDMA systems*, GLOBECOM 2006.
[4] Xing Z., En Z., Renshui Z., Shiming L., and Wenbo W.: *Adaptive Multiuser Radio Resource Allocation for OFDMA Systems*, GLOBECOM 2005, Vol.6, Dec. 2005.
[5] Xing Z., and Wenbo W.: *Multiuser frequency-time domain radio resource allocation in downlink OFDM systems: Capacity analysis and scheduling methods*, Computers and Electrical Engineering, Vol.32 Issue.1-3, pp. 118–134, 2006.
[6] Sang S. J., Dong G. J., and Wha S. J.: *Cross-layer Design of Packet Scheduling and Resource Allocation in OFDMA Wireless Multimedia Networks*, VTC 2006, Vol.1, pp. 309–313, 2006.
[7] Seungwan R., Byunghan R., Hyunhwa S., and Mooyong S.: *Urgency and Efficiency based Packet Scheduling Algorithm for OFDMA wireless System*, ICC 2005, Vol.5, pp. 2779–2785, May. 2005.
[8] Ahmed K. F., and Khaled M. F.: *Opportunistic Scheduling of Delay Sensitive Traffic in OFDMA-baesd Wireless Networks*, WoWMoM 2006, Vol.2006, pp. 279–288, June 2006.
[9] Sanjay S., and Alexander L.S.: *Scheduling Algorithms for a Mixture of Real-Time and Non-Real-Time Data in HDR*, Proceedings of International Teletraffic Congress (ITC), 2001.