Opportunistic Scheduling of Delay Sensitive Traffic in OFDMA-based Wireless Networks

Ahmed K. F. Khattab,* Student Member IEEE
Rice University
Dept. of ECE, Houston, TX 77005
khattab@rice.edu

Khaled M. F. Elsayed, Senior Member IEEE
Cairo University
Dept. of ECE, Giza, Egypt 12613
khaled@ieee.org

Abstract

OFDMA is an attractive multiple access technique for packet-based mobile broadband wireless access for beyond 3G and 4G systems. Radio resource allocation in OFDMA can exploit multiuser diversity to increase system capacity by implementing opportunistic scheduling techniques. This paper presents a new opportunistic scheduling scheme for OFDMA-based wireless multimedia networks. We focus the scheduling algorithm on the class of delay-sensitive packets that belong to interactive applications such as telephony and video streaming. We divide the scheduling decision into two sub-problems: the OFDMA subcarrier allocation and subsequently the subcarrier assignment. Both the subcarrier allocation and assignment algorithms exploit multiuser diversity and are designed to provide fairness with respect to the realizable throughput per user, packet dropping ratios and packet delay distributions. We investigate various performance aspects of the proposed scheduling algorithm using actual MPEG-4 traffic traces under different system loading and requested deadline values. The results show the superiority of the proposed scheduling scheme and its excellent performance with respect to throughput, packet dropping, and delay distributions.

1. Introduction

The increasing demand for high transmission rates in wireless communication systems brought into play the need for new transmission techniques. However, high transmission rates may result in sever frequency selective fading and intersymbol interference (ISI). Orthogonal Frequency Division Multiplexing (OFDM) has recently been proposed as an effective multi-carrier solution for broadband wireless transmission [1]. In an OFDM system, the high data rate stream is transformed into a number of lower rate components. Each of the OFDM signal components is modulated onto a distinct subcarrier. The bandwidth of the low rate component is narrower than the coherence bandwidth of the channel, thus the transmission in each subcarrier experiences flat fading. Furthermore, orthogonal subcarriers cause OFDM systems to have a higher spectral efficiency. These advantages made OFDM to be adopted for the physical layer in many current and future high speed wireless communications systems such wireless local area networks (WLAN), wireless metropolitan area networks (WMAN), and mobile broadband wireless access (MBWA) standards.

In multiuser environment, OFDM can also be applied producing a highly flexible, efficient high speed communications system. Since Wahlqvist *et al* [2] studied multiuser OFDM, intense research was carried out aiming to find improved and flexible multiple access methods other than traditional time division multiple access (TDMA) or frequency division multiple access (FDMA) techniques (which employ fixed and predetermined time-slot or subcarrier allocation schemes). Orthogonal frequency division multiple access (OFDMA) is a promising multiple access scheme that has recently attracted enormous research interest [4]-[13]. (OFDMA is a multiple access scheme which is based on OFDM with the only exception that the OFDM symbol is composed of data from multiple users sharing the wireless system.) Here, the base station is responsible for deciding how the available subcarriers will be distributed among different users.

In this paper, we consider the subcarrier management problem in the downlink of OFDMA wireless multimedia networks for delay-constrained traffic. The problem is divided into two sub-problems: the subcarrier allocation problem and the subcarrier assignment problem. Based on the principles of multi-user diversity [14], we propose an opportunistic subcarrier allocation algorithm that uses the channel state information and the delay information of different downlink flows to calculate the number of subcarriers to be assigned to each active user in the system. The algorithm also attempts to guarantee the QoS required by

*This work was done when the author was at Cairo University.

these users. We also propose an opportunistic algorithm for the subcarrier assignment problem. The proposed algorithm monitors the deadline violations in all queues, and ensures fairness among different users in their service rates. This is achieved by distributing the deadline violation occurrence among all flows evenly.

The rest of this paper is organized as follows: Section 2 defines the OFDMA network model. Then we describe the multiuser scheduling problem in such networks in section 3 and include a survey of related work. In section 4, the proposed subcarrier allocation and subcarrier assignment algorithm are introduced. We report the results of extensive set of simulation experiments in section 5. Section 6 summarizes the main findings of the paper.

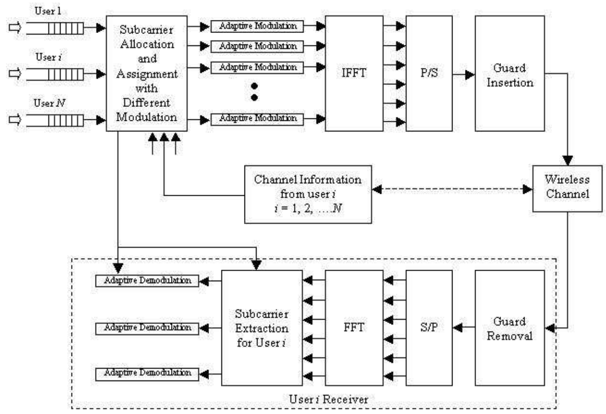

Figure 1. OFDMA transceiver architecture.

2. The OFDMA Network Model

We consider the downlink scheduling of a single cell in cell-structured OFDMA-based system. The cell is equipped with a base station which is responsible for coordinating the simultaneous transmissions of N mobile users over S OFDM subcarriers. Inter-cell interference is not taken into consideration. OFDMA adds multiple access to OFDM by allowing a number of users to share an OFDM symbol. Therefore, an OFDMA transmitter employs a subcarrier allocation and assignment function instead of the serial to parallel conversion used in OFDM systems to split the single user's stream into a set a parallel low rate streams. The rest of OFDMA system is the same as an OFDM system as shown in Figure 1. Adaptive modulation is used to transmit data over individual subcarriers with different gains yielding significant performance improvement. The base station is responsible for informing each user terminal which subcarriers are assigned to it via a set of subcarriers (or time slots in a frame) reserved for control functions. The receiver then does the reverse operations at the transmitter and the data sent to this user is retrieved by demodulation of the user's assigned subcarriers.

High speed wireless standards are usually operated at high frequencies. High frequency channels, like those used in OFDMA-based networks, are characterized by their time-varying, frequency-selective fading nature. Channel gains vary from subcarrier to subcarrier for a single wireless terminal due to multipath propagation. Besides, channel gains of each subcarrier vary over time for the same user terminal, due to the movement of the terminal and other objects within the surrounding area. At a given time, some subcarriers suffer severe fading, while others have a good response. In this case, if the channel information is available, it is more efficient to transmit data only over those subcarriers having a good response with high transmission rates. Furthermore, channel gains of a specific subcarrier vary from wireless terminal to wireless terminal due to statistical independence. This implies that certain subcarriers that are in deep fade for some users are not necessarily bad for others since the user channel fading characteristics are uncorrelated for different users. Hence, selection of good subcarriers for one user may not necessarily block other users from accessing their good subcarriers. This gives the general motivation to develop a resource allocation framework that exploits multiuser diversity to allocate and assign an active user its best subcarriers, and hence increase the efficiency of channel utilization.

3. Scheduling in OFDMA Networks

In OFDMA-based networks, the scheduler is responsible for dividing the set of subcarrier available to the base station into a number of mutually disjoint subsets of subcarriers. Each subset is assigned to a certain user for a certain period of time (scheduling interval). Recently, there has been intensive research on subcarrier and bit allocation in multiuser OFDMA systems [4]-[13]. Those algorithms can be categorized as static and dynamic allocation algorithms.

3.1. Static Subcarrier Management

Static subcarrier management schemes depend on traditional multiple access schemes, such as Time Division Multiple Access (TDMA) and Frequency Division Multiple Access (FDMA), as a mechanism for distributing the subcarriers in multiuser OFDM networks. In OFDM-TDMA, one of the users is assigned all the subcarriers for the entire scheduling interval, whereas in OFDM-FDMA, each user is permanently assigned one or several predetermined subcarriers. Thus, both OFDM-TDMA and OFDM-FDMA are not capable of adapting to the channel gain variations. As a consequence, any fixed assignment of subcarriers to terminals will waste system resources in the form of either power or bit rate [3],[5].

3.2. Dynamic Subcarrier Management

Recently, dynamic radio resource management schemes that consider the users' instantaneous channel conditions have attracted enormous research interest. This is due to the significant overall system efficiency increase obtained when variations in channel gains among users (multiuser diversity) is exploited. These schemes vary in their design and performance objectives, however, they can generally be classified into two main categories: the first set of schemes target the minimization of total power subject to a minimum achievable throughput (total or per user), while the second set of schemes target the maximization of throughput subject to maximum power (total or per user).

Many schemes known as bit loading algorithms have been suggested [3]-[6]. They adapt transmission power or bit rates optimally to the channel gains of different subcarriers, where either a feasible overall bit rate or a maximum available transmit power is given. They are based on a result from information theory describing how to distribute transmission power over a set of subcarriers with different channel gains in order to maximize the channel's capacity. This is known as the water filling principle first discussed by Shannon [15].

Wong *et al* [3] addressed the problem of minimizing the transmitted power at a given bit rate per terminal. Subcarrier and bit allocation was done dynamically through the use of nonlinear optimization with integer variables. A modification to Wong's algorithm was proposed in [4]. The proposed extension allows each user to specify its individual QoS requirements, defined in terms of bit rate and bit error rate. The scheme distributes subcarriers and transmit power among multiple users according to their QoS requirements.

An alternative problem of maximizing the overall bit rate of multiple wireless terminals while the transmit power is upper bounded have been considered in [6] by Yin and Liu. Such a problem can be decomposed into two tasks: first determining the number of subcarriers each terminal is assigned and then choosing which subcarriers to be assigned. The subcarriers assignment problem is solved by mapping to the maximum weight perfect matching problem. While this algorithm computes the optimal assignment of subcarriers to terminals, it is still fairly computationally expensive.

Due to the high complexity of the optimization techniques used in the above work, many suboptimal heuristic algorithms have been proposed [7]-[13]. The computational complexity of the heuristics is considerably smaller than that of the optimum solution. On the other hand, their results are usually close to the optimum solutions [7]. For example, an iterative solution of the total transmit power minimization problem is proposed in [7]. Subcarrier allocation with fixed modulation is used and then the bit loading scheme is used to maximize the total number of transmitted bits. The authors also introduce a resource allocation scheme whose objective is to maximize capacity, based on the proportional fair scheduling algorithm for point-to-point communication.

A dynamic subcarrier and bit allocation algorithm, which takes advantage of the knowledge of instantaneous channel gain in subcarrier and bit allocation, is presented [8]. Initially, the greedy single user water-filling approach is used to allocate subcarriers and bits as if all the subcarriers in the system could be used exclusively by this user. In case that a subcarrier is desired by more than one user, the algorithm will arbitrate the subcarrier to one user appropriately so that total transmit power is minimized.

In [9], a scheduling scheme for increasing the number of non-real-time users with a minimum bit rate requirement is proposed. Svedman *et al.* [10] present a QoS-aware proportional fair scheduler whose objective is to mainly increase the overall system throughput while providing fairness among the active traffic streams. The scheme can be integrated with an opportunistic beamformer to increase system fairness. The scheme is evaluated for a densely populated system with a large number of subcarriers. Also, subcarrier clustering is used to reduce the amount of signaling information needed for the exchange of subcarrier gains from the terminals to the access point.

When other types of QoS requirements, such as the delay requirements of real-time traffic, are to be considered, the problem at hand becomes much more complicated. This is pursued in [11]-[13]. The authors of [11] and [13] follow the Yin's [6] decomposition approach and present two related heuristic algorithms for the problem of assigning subcarriers to terminals, assuming that the number of subcarriers allocated to each terminal is fixed [11]. These algorithms assign each user the current best subcarriers in a prioritized manner. Then they present a new subcarrier allocation method that finds the number of subcarriers to be assigned to each user in the start of every time slot [12]. The method is based on allocating subcarriers for terminals depending on the actual queue size of each terminal relative to the overall data queued at the access point. In contrast to other methods described above, channel gain information as well as any flow specific knowledge is not explicitly included in the allocation of subcarriers, however this kind of information is indirectly reflected in the queue size.

The proposed adaptive resource allocation algorithm in [13] is based on the channel conditions and power limitation observed in the physical layer, as well as the queue status, packet arrival, QoS requirements, service discipline, and user fairness observed at the data link layer. The objective is to minimize the overall transmission power while maintaining the channel errors at a sufficiently low level, so that the assumption of error-free channel in the scheduling part is generally valid. With error-free links, the system can fairly guarantee various QoS requirements to all the users from the physical-layer's point of view.

The work presented here is mostly related to [11]-[13], however it offers several advantages since no explicit provisioning for delays or fairness is considered in the formulations in [11]-[13]. Our work distinguishes itself by taking the delays (which also carries queue size information) into the formulation and by setting the objective to achieving the maximum throughput in the downlink (where transmit power is not severely limited). Fairness and multiuser diversity are explicitly targeted in the subcarrier assignment and allocation schemes.

4. The Proposed Scheduling Scheme

We introduce a new opportunistic approach of subcarrier management in OFDMA-based wireless multimedia networks. The idea behind this approach is that subcarrier allocation and assignment is not only dependent on the instantaneous channel conditions of different users, but also on the QoS requirements and fairness among users. The QoS requirement of real-time traffic is usually defined in terms of a delay bound before which the packet should be delivered to the receiver. Otherwise, the information contained in this packet will be of no use to the receiver. Before we formulate our problem let us define the following parameters:

N : The number of users in the system (each user is assign a separate queue).

N_t : The number of active users (with at least one valid packet queued) at time t.

S : The total number of data subcarriers available to the system.

$n_i(t)$: The number of subcarriers to be assigned to the i^{th} user at the slot starting at time t.

r_i : The average traffic rate of i^{th} user.

$\mu_{ij}(t)$: The channel capacity (the maximum possible data transmission rate) of the subcarrier number j if allocated to the i^{th} user.

$\bar{\mu}_i(t)$: The average subcarrier capacity of the i^{th} user (if it was allocated all the subcarriers).

$$\bar{\mu}_i(t) = \frac{1}{S}\sum_{j=1}^{S}\mu_{ij}(t) \tag{1}$$

$d_i(t)$: The time to expire of the i^{th} user head of line (HoL) packet, which is the difference between the deadline, T_i, and the HoL packet delay up till time t, $W_i(t)$, i.e

$$d_i(t) = T_i - W_i(t) \tag{2}$$

$V_i(t)$: The number of deadline due violations of the i^{th} user packets up to time t.

$\delta_{ij}(t)$: An indicator of allocating the subcarrier j to the user i in the time slot starting at time t.

$$\delta_{ij}(t) = \begin{cases} 1 & \text{if subcarrier } j \text{ is assigned to user } i \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

We assume that perfect channel information is known at both the transmitter and the receiver. Mobile terminals may be equipped with mechanisms to measure the rate at which they have been served. This data may be reported back to the base station, so that it can estimate the channel of all mobile channels based on that data as long as the channel variation is slow. As a result, the resource allocation should be done within the coherence time of the channel.

Our objective of the resource allocation problem can be defined as maximizing the total system throughput subject to a QoS constraint that the HoL packet delay $W_i(t)$for active users is less than a given value T_i. Let's define the total instantaneous system throughput $R_T(t)$ as the aggregate traffic transmitted over the system at time t. Thus,

$$R_T(t) = \sum_{i=1}^{N}\sum_{j=1}^{S}\delta_{ij}(t)\mu_{ij}(t) \tag{4}$$

The subcarrier management problem is formulated as follows:

$$\max_{\delta_{ij}(t)} R_T(t) \qquad for\delta_{ij} \in \{0,1\} \tag{5}$$

subject to

$$\sum_{i=1}^{N}\sum_{j=1}^{S}\delta_{ij}(t) = S \tag{6}$$

and

$$W_i(t) < T_i \qquad i = 1, 2, ..., N \tag{7}$$

Note that constraint (7) also includes to some extent the history of the previous assignments. When the subcarrier assignments of a certain user during the previous slots do not satisfy the delay requirement, its waiting time approaches its deadline. Constraint (7) attempts to force the current assignments to compensate such user by either allocating more subcarriers or assigning higher quality subcarriers in order to prevent its packets from expiry. However, a mathematical expression that formulate the relationship between the waiting time of the HoL packet of a certain queue and the subcarriers assigned to it (either in the current assignments or the previous assignments) can not be explicitly evaluated without restrictive assumptions on the arrival process and the time-varying subcarriers response. Thus, a mathematically optimal solution of the above problem in its most generic form cannot be obtained.

If this relationship between successive subcarrier assignments and HoL packets delays could be found, the above problem can be solved with Integer Programming (IP). We refer to this approach as the optimal solution to the resource allocation problem. Although the optimal solution gives the exact results, from an implementation point of view, it is not preferred. In a time varying channel, it is required to allocate the subcarriers within the coherence time. In most non-trivial cases (large number of users, large number of subcarriers), this cannot be achieved by the NP-hard IP solution. This real-time implementation requirement leads to the quest for suboptimal solutions that are fast and close to the optimal solution.

In what follows we propose a heuristic suboptimal solution of the above problem. We follow the decoupling approach of this NP-hard problem proposed by Yin and Liu [6]. Our solution is done in two steps:

1) Subcarrier Allocation: deciding **how many** subcarriers to be assigned to each user (i.e. determines $n_i(t)$).

2) Subcarrier Assignment: determining **which** subcarriers to be assigned to each terminal (i.e. the vectors $\delta_{ij}(t)$ are calculated).

4.1. Subcarrier Allocation Algorithm

We first determine the number of subcarriers $n_i(t)$ to be assigned to every user in the set of active users N_t at that time instant. Our allocation is based on three factors: 1) the instantaneous subcarrier channel gains of active users, 2) the users' average rates, and 3) the delay of the HoL packets of these users. We not only exploit the statistical variations of the users' channels, but also use the statistical variations of users' queues in order to increase the efficiency of channel throughput utilization. In what follows we explain the three steps of the proposed subcarrier allocation algorithm illustrated in Figure 2.

Step 1: Initially each active user is allocated a number of subcarriers $\acute{n}_i(t)$ given by:

$$\acute{n}_i(t) = \left\lceil \frac{r_i}{\frac{1}{|N_t|}\sum_{j\in N_t} r_j} \frac{\bar{\mu}_i(t)}{\frac{1}{|N_t|}\sum_{j\in N_t} \bar{\mu}_j(t)} \right\rceil \qquad (8)$$

In essence, this allocation exploits multiuser diversity by allocating more subcarrier to the users with better channels. For instance, lets us assume that the average traffic rate of all users is the same, then the factor $r_i/\frac{1}{|N_t|}\sum_{j\in N_t} r_j$ is equal to one. A user with relatively good channel conditions, i.e. its $\bar{\mu}_i(t) > \sum_{j\in N_t} \bar{\mu}_j(t)/\mid N_t \mid$, will initially be allocated two or more subcarriers. On the other hand, a user with relatively bad channel conditions, will initially be allocated only one subcarrier. The role of the weighting factor $r_i/\frac{1}{|N_t|}\sum_{j\in N_t} r_j$ is to weigh the allocation proportional to users' average rates.

At the end of this step, if all the available data subcarriers are allocated to the set of users currently seeking service

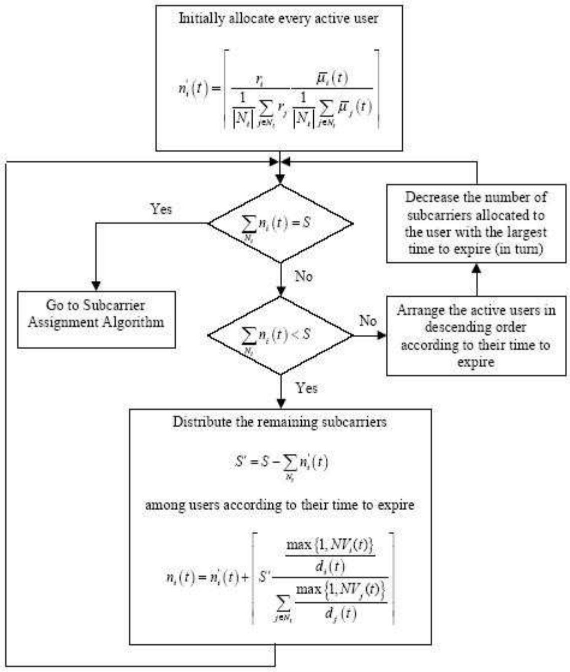

Figure 2. Subcarrier Allocation Algorithm.

from the system, the allocation algorithm terminates. However, if some subcarriers remain unused after this step, the unused subcarriers are allocated to some of the active users. Let us denote the number of the remaining unused subcarriers by $\acute{S}$, where

$$\acute{S} = S - \sum_{N_t} \acute{n}_i(t) \qquad (9)$$

The next two steps of the algorithm are responsible for distributing these remaining subcarriers efficiently among the active users in order to prevent as many packets from expiry, and thus being dropped by the system.

Step 2: In order to reduce packet droppings, the second step of the algorithm should allocate the biggest share of the remaining subcarriers $\acute{S}$ to the user with the smallest time to expire $d_i(t)$. Since the value of $1/d_i(t)$ increases significantly as the time to expire $d_i(t)$ decreases, the proposed algorithm will distribute the remaining set of subcarriers S' among the active users according to the ratios of $1/d_i(t)$. When the system is heavily loaded with users, deadline due violations start to occur. The proposed allocation algorithm adapts to this situation by considering the number of violations $V_i(t)$ in calculating the share of additional subcarriers. Using $V_i(t)/d_i(t)$ as the distributing ratio, users recently suffering from more violations (than the average of all users) will be compensated by allocating them more subcarriers. The number of violations reflects the history of the previous assignments within a certain time window (we take the window size equal to about 1000 scheduling intervals). This step enhances fairness in distributing the violations occurrences among all users. At this point, the number of subcarriers to be assigned to the i^{th} active user is now

$$n_i(t) = \acute{n}_i(t) + \left\lceil \acute{S} \frac{\frac{\max\{1, V_i(t)\}}{d_i(t)}}{\sum_{j \in N_t} \frac{\max\{1, V_j(t)\}}{d_j(t)}} \right\rceil \qquad (10)$$

If the total number of allocated subcarriers equals the available number, the algorithm terminates, and then moves to the subcarrier assignment algorithm. However, the second component of the equation (10) may cause the total number of allocated subcarriers $\sum_{j \in N_t} n_j(t)$to be greater than the available subcarriers S. In the next step of our proposed resource allocation algorithm, we ensure that the number of subcarriers allocated is exactly equal to those available in the system.

Step 3: The last step in our subcarrier allocation procedure is used (if needed) to decrease the number of subcarriers allocated to some users so that the total allocated subcarriers equals S. Our criteria in choosing these users, whose number of allocated subcarriers are to be decreased, is the time to expire of their HoL packets and their violation occurrences similar to what was done in step 2 of the algorithm. First, the algorithm sorts the set of active users in a descending order according to their HoL packet time to expire. Then, it iterates over users in that order. In every iteration, the algorithm decreases the number of subcarriers allocated to the user in turn by one. It then checks whether the total subcarrier allocated equal to S or not. If it was not yet equal, the algorithm iterates once more.

In this step of the algorithm, users how have more persistent delay requirements (their HoL packets are to expire) are allowed to keep their previously allocated subcarriers. While users whose packets' are far from expiry may lose one or more of their allocated subcarriers. Thus, some active users could have zero allocated subcarriers in certain scheduling intervals. Such a procedure will enhance the overall system performance and capacity as will be demonstrated later via simulation.

The computational complexity of this algorithm is at most $O(N_t log(N_t))$. As in step 3, the active users in the system may be sorted once according to their time to expire once (we remind that step 3 is optional thus the actual complexity is actually less than $O(N_t log(N_t))$).

4.2. Subcarrier Assignment Algorithm

The objective of the subcarrier assignment algorithm is to find the subcarrier assignment that maximizes the total rate. This can be achieved if multiuser diversity is used to assign every active user its best $n_i(t)$ subcarriers. Such assignment problem is equivalent to the maximum weighted perfect matching problem in bipartite graphs [8]. An optimal solution can be generated by the Hungarian algorithm [16], which has the complexity of $O(S^3)$, where S is the number of subcarriers.

With the objective of enhancing the fairness characteristics of the scheduling algorithm while maximizing the total rate, we propose a new low complexity opportunistic subcarrier assignment algorithm. The proposed algorithm is a priority based algorithm that favors users with more packect droppings. At some time instant (the start of a time window), the algorithm assigns initially a unity priority to all users. Whenever a packet is dropped from a certain user's queue, that queue priority is incremented by one. In every scheduling interval, the subcarrier assignment algorithm sorts the active users in the system in a descending order according to their priorities. The user with the highest priority is allowed to pick those subcarriers with best channel response from the set of all subcarriers. After assigning those subcarriers to that user, the algorithm removes them from the set of available subcarriers. Then the algorithm assigns the next higher priority user the best set of remaining subcarriers, and so on. This mechanism should enhance the fairness performance of the scheduler with respect to deadline violations, and consequently rates (since rates are dependent on dropping ratios). If more than one user share the same priority level, ties are broken by giving priority to the user with the best channel quality (averaged over all subcarriers) to pick up its allocated subcarriers first. Hence, the proposed assignment algorithm also exploits multiuser diversity in assigning the available subcarriers to the active user in the system (by letting the user with the best channel conditions at ant time instant pick its allocated subcarriers before the next better channel user, and so on). Figure 3 summarizes the proposed subcarrier assignment algorithm. The algorithm assigns S subcarriers to N_t wireless terminals. Each user picks up its allocated subcarriers from a sorted list, which has to be regenerated for each user (due to removal of some subcarriers assigned to the previous user). Thus, the worst-case computational complexity of this assignment procedure is given by $O(N_t S log(S))$. Even when added to the complexity of sorting the active users according to their priorities $O(N_t log(S))$, the computational complexity of complete subcarrier assignment algorithm is much lower than that of the optimal (Hungarian) algorithm.

In the next section, we carry out an extensive set of simulation experiments in order to gain insight of the different performance aspects of the proposed subcarrier allocation and assignment algorithms. Specifically, we are interested in evaluating the throughput, delay, and fairness characteristics of the algorithms.

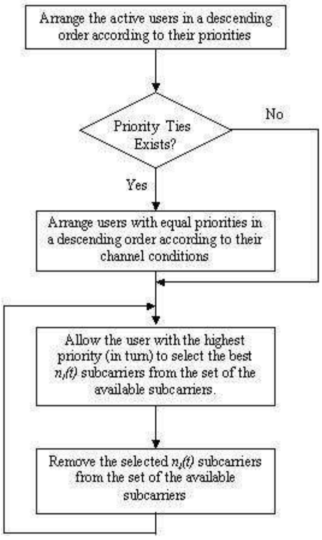

Figure 3. Subcarrier Assignment Algorithm.

5. Performance Evaluation

5.1. Simulation Setup

We consider the IEEE 802.20 Mobile Broadband Wireless Access (MBWA) system model [17]. MBWA systems are to be designed to provide a broadband, IP-oriented connection to a wireless user that is comparable to wired broadband connections that are in use today [18]. Among the different channel bandwidths suggested in [19], we use a channel of 5 MHz bandwidth. We assume that the number of data subcarriers S is equal to 128 subcarriers. Adaptive modulation is used to transmit data on each subcarrier such that the highest possible rate can be transmitted in every scheduling interval. The scheduler chooses the modulation order out of five modulation types available (BPSK, QPSK, 16-QAM, 64-QAM and 256-QAM). Traffic is generated from a trace file of a 30 frames/sec MPEG-4 encoder with an average rate of 256 Kbps and a peak rate of 2.3 Mbps [21]. Each frame is decomposed into 50 bytes packets to be transmitted. A packet is assigned a deadline before which it should be served. We consider 100, and 500 milliseconds delay bounds. Such values are reasonable for video streams. The simulation environment was implemented using the IT++ package [20] with the values listed in Table 1. The simulation period of the following experiments is 30 seconds.

5.2. Results and Discussions

In this section we present the numerical results of our opportunistic subcarrier management algorithms. For comparison reasons the results of the dynamic scheme proposed by Gross *et al.* in [11] and [12] (we call it here the GKKW algorithm) are included. As discussed in section 3, this algorithm is the closest subcarrier management scheme to our algorithm as it first determine the number of subcarriers to be assigned to the i^{th} terminal at a time instant $t : n_i(t)$ as

$$n_i(t) = 1 + \left\lfloor (S - N_t)\frac{L_i(t)}{\sum_{\forall j} L_j(t)} + 0.5 \right\rfloor \quad (11)$$

where $L_i(t)$ is the length of the queue of the i^{th} user. Then, the GKKW algorithm assigns each user the current best subcarriers in a prioritized manner (determined in a circular order). It worth mentioning that no other work in the literature achieves better performance than this algorithm.

Moreover, we are including the simulation results of using OFDM-TDMA subcarrier assignment with multiuser diversity scheduling based on the Exponential rule [22] scheduling discipline. This scheme is referred to here as OFDM-TDMA/EXP. It is chosen since it is reported in the literature as the best TDMA-based discipline for scheduling real-time traffic over time varying channels. This scheme schedules the j^{th} user at the time slot starting at time t for transmission, where

$$j = \arg\max_i \{a_i \frac{\mu_i(t)}{\tilde{\mu}_i(t)} e^{\frac{a_i W_i(t) - \overline{aW}}{1+\sqrt{\overline{aW}}}}\} \quad (12)$$

where $\mu_i(t)$ is the total channel capacity of the i^{th} user (i.e. $\sum_{j=1}^{S} \mu_{ij}(t)$), and $\tilde{\mu}_i(t)$ is the exponentially smoothed average of $\mu_i(t)$. Also, $a_i = -log(\delta_i)/T_i$, with δ_i being the maximum probability (taken here as 5%) of $W_i(t)$ exceeding T_i, reflect the QoS requirements of the real-time users. We assume that all users belong to the same service class with the same QoS requirement and thus, T_i and δ_i are the same for all users. Both the GKKW algorithm and the OFDM-TDMA/EXP assignment scheme are used as a benchmark against which the performance gains achieved by our opportunistic subcarrier allocation and assignment are illustrated.

Carrier frequency	1.9 GHz
Channel bandwidth	5 MHz
Number of data subcarriers	128
User mobility speeds	3 - 120 Km/hr
Doppler frequency	211 Hz
Doppler spectrum	Jakes' (6 rays)
Scheduling interval	1.667 msec
Traffic model	Real trace file of MPEG-4 encoder Frame rate: 30 frames/sec
Traffic average / peak rate	256 Kbps / 2.3 Mbps
Packet size	50 bytes

Table 1. Summary of simulation parameters.

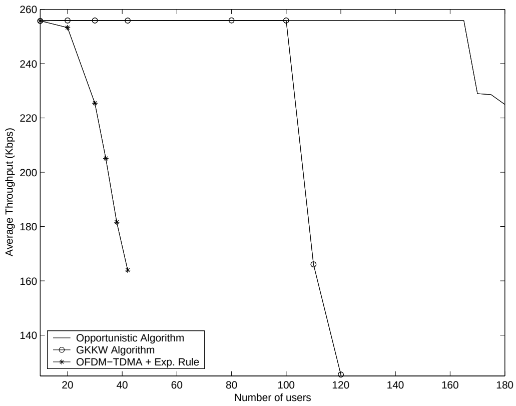

Figure 4. The average throughput per user for 100 msec delay bounds.

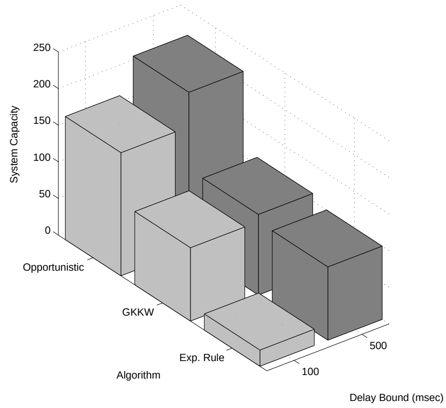

Figure 5. System Capacity for different algorithms and delay bounds.

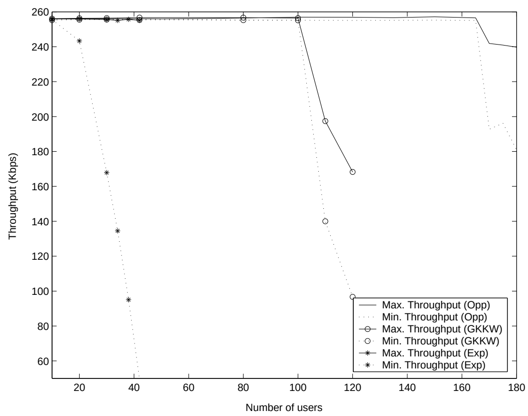

Figure 6. The maximum and the minimum achieved for 100 msec delay bounds.

1) Capacity and Throughput

Firstly, we find out the number of users that the system can serve without degrading the service offered to them. Defining the system capacity as the number of users beyond which the average throughput per user falls to 98% of the average arrival rate. We find that when the opportunistic allocation/assignment is used, 215 and 168 users can be served by the system for a 100, and 500 milliseconds delay bound, respectively. While only 110 and 100 users could be handled if the GKKW algorithm is used for the same delay bound, respectively. Figure 4 depicts the average throughput per user versus the number of users for different algorithms for a 100 milliseconds delay bound. It worth mentioning that significant gain in the system capacity offered by our proposed algorithm over the GKKW algorithm is owed to the dynamicity of our subcarrier allocation algorithm. Our proposed subcarrier allocation algorithm may not allocate any subcarriers to some of the active users if their channel conditions are bad or their delays are far from the delay bound. The unused subcarriers are allocated to users with better channel conditions or to users with packets approaching the deadline expiry bound. On the other hand, the GKKW allocation necessitates that every active user should have at least one subcarrier as clear from (11), which limits the maximum number of users to the number of available subcarriers S (128 in our case).

For the case of OFDM-TDMA/EXP, the system capacity falls to only 22 and 100 users for the 100, and 500 milliseconds delay bounds, respectively. This is due the inefficiency caused by the assignment of all subcarriers to one particular user at a certain scheduling interval. As we discussed earlier, while some subcarriers are suffering bad conditions for a certain user they may be the best subcarriers for other users. Moreover, with strong delay requirements (such as the 100 milliseconds bound), the system capacity is severely degraded due to the lack of an efficient dynamic mechanism to compensate users approach deadline due violation, and therefore a huge amount of traffic is dropped by the system. Figure 5 summarizes the system capacities for different algorithms and delay bounds. Thus, the proposed subcarrier allocation and assignment algorithm can be the best choice for application in future MBWA systems.

2) Fairness

In order to qualify the fairness characteristics of the proposed scheduling scheme, we show the maximum and the minimum value of the users' throughput for the above experiments in Figure 6. It is noticed that the difference between the maximum and the minimum achieved throughput is insignificant when the system is serving a number of users less than or equal to the system capacity (defined earlier) using either the proposed opportunistic algorithm or the GKKW algorithm. However, when the number of users

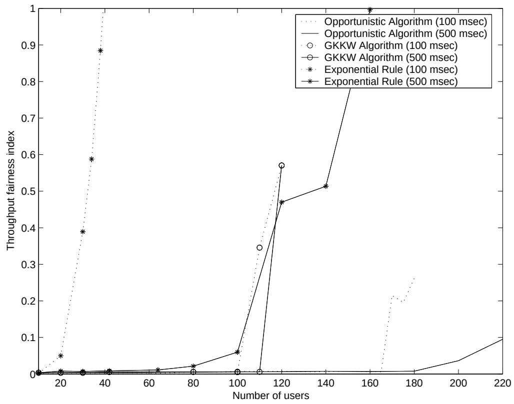

Figure 7. Throughput fairness indices of different algorithms and delay bounds.

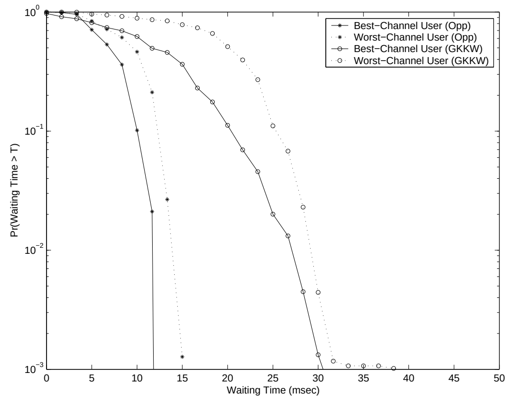

Figure 8. Delay tails of users with the best and worst channels for 500 msec bound.

is larger than the system capacity, the difference between the maximum and the minimum achieved throughput using the proposed algorithm is much less than that achieved by the GKKW algorithm. This superior performance is since our algorithm attempts to minimize the loss from the active user queues by allocating more subcarriers to those who have experienced recent packet loss. Even when the system is operated with users more than its capacity, the fairness of distributing these lost packets among all users is achieved via the subcarrier assignment algorithm. On the other hand, for the OFDM-TDMA/EXP assignment, that difference is highly noticed and is dependent on the value of the delay bound. Again, the reason of this behavior is the lack of efficient compensation mechanism for users with bad channels. In order to formalize the throughput fairness performance, we define the throughput fairness index as the ratio of the difference between the maximum and the minimum achieved throughput (λ_{max} and λ_{min}, respectively) to the average throughput per user (λ_{avg}), i.e.

$$ThroughputFairnessIndex = \frac{\mid \lambda_{max} - \lambda_{min} \mid}{\lambda_{avg}} \quad (13)$$

The fairness indices of all scheduling algorithms and delay bounds are plotted in Figure 7. The proposed algorithm exhibits almost-perfect fairness regardless the value of the delay bound (fairness index approaches 0). OFDM-TDMA/EXP subcarrier assignment lacks such fairness.

3) Delay Performance

Next, we inspect the delay performance of the proposed subcarrier allocation and assignment algorithms. This is achieved through investigating the distribution of the delays that users' packets incur at the base station. A good scheduling algorithm should keep all delays below the delay bound with high probability. Due to the large number of users the system can support, we focus only on the delay distributions of the user with the best channel quality and the user with the worst channel quality. The delay distributions of these particular users are sufficient to evaluate the delay performance of the scheduling algorithm since they represent the opposite extremes of the channel qualities. The delay distributions of the best and the worst channel users for a 500 milliseconds delay bound with 100 users are shown in Figure 8. The delay tails of OFDM-TDMA/EXP are very close to the deadline, while our opportunistic scheme and the GKKW algorithm keep the delays of all users far below the deadline (few tens of milliseconds) and close to each other. Moreover, 5.3% of the worst channel user's packets were lost using the OFDM-TDMA/EXP scheduling. So we don't show the results of the OFDM-TDMA/EXP. Another observation is that the delay distributions and the maximum delays of the best and the worst channel users of our opportunistic algorithm is slightly better than those of the GKKW algorithm (this remarkably were consistent for other values of the delay bound not shown here due to space limitation).

The good delay performance of the opportunistic algorithm comes from the used mechanism of compensation of low-quality users. When some users are suffering from a long period of bad channel conditions, their HoL packet delays increase, so that the allocation algorithm allocates more subcarriers to these users (even on the expense of taking a part or all of the subcarrier initially assigned to other higher-quality users). Preventing high-quality users from their subcarrier shares increases their HoL packets delay. Thus in the next scheduling interval the algorithm allocates them more subcarriers, and so on. Hence, the delays of all users are always kept small and below the deadline. It is worth mentioning that this mechanism is not affected by the value of the delay bound since it is only concerned by the difference in HoL delays of different users. Even with high user population (not shown here for space limitations), the opportunistic subcarrier assignment and allocation algorithms managed to keep the delays of users with different channel conditions very close.

Though rather a computationally inexpensive algorithm, the proposed opportunistic scheduling algorithm can be used to provide statistical delay guarantees for time-sensitive traffic in OFDMA-based wireless networks. The algorithm exhibits a unique fairness behavior in the services (packet delays and throughput) delivered to different users. Moreover, multiuser diversity is used in the algorithm to offer orders of magnitude increase in the system capacity.

6. Conclusion

This paper addresses the problem of scheduling real-time users over OFDM-based wireless multimedia networks. We introduced new opportunistic subcarrier allocation and assignment mechanisms for parallel transmission of data streams to different terminals in OFDMA-based broadband wireless systems. The subcarrier allocation algorithm instantaneously determines the number of subcarriers each terminal should receive by the assignment algorithm for the next downlink scheduling interval. Gains in throughput and realized delay are achieved by exploiting multi-user diversity techniques, the subcarrier allocation algorithm takes into account the current channel state for each user in the system, as well as other stream specific delay information (the time to expire of the HoL packet) and the number of recent deadline violations. The allocated number of subcarriers is assigned to terminals dynamically in a manner that ensures fairness in the deadline violation occurrences among different users.

The proposed algorithms outperforms existing ones in the sense that the services received by different real-time users, namely, delays, rates, and loss ratios, are fairly achieved for a wide of applications and user population. The proposed policies have low computational complexity and are suitable for application in future broadband wireless systems such as the IEEE 802.16a and the 802.20 MBWA systems. Future extensions should tackle multi-cell environment, uplink scheduling, and practical consideration for channel state information signaling or exploiting estimation techniques to reduce signaling overhead.

References

[1] R. Nee, and R. Prasad. *OFDM for wireless multimedia communications.* Artech House universal personal communications library, 2000.

[2] M. Wahlqvist, C. stberg, J. Beek, O. Edfors, P. Brjesson. A Conceptual Study of OFDM-based Multiple Acess Schemes. *Technical Report Tdoc 117/96, ETSI STC SMG2 meeting no 18*, Helsinki, Finland, May 1996.

[3] C. Wong, R. Cheng, K. Letaief, and R. Murch. Multiuser OFDM with adaptive subcarrier, bit and power allocation. *IEEE Journal on Selected Areas of Communications*, 17(10):1747-1758, October 1999.

[4] S. Pietrzyk and G. Janssen. Multiuser subcarrier allocation for QoS provision in the OFDMA systems. *Proc. of Vehicular Technology Conference (VTC)*, Birmingham, May 2002.

[5] W. Rhee and J. Cioffi. Increase in capacity of multiuser OFDM system using dynamic subchannel allocation. *Proc. Vehicular Technology Conference (VTC)*, Boston, May 2000.

[6] H. Yin and H. Liu. An efficient multiuser loading algorithm for OFDM-based broadband wireless systems. *Proc. IEEE Globecom*, 2000.

[7] M. Ergen, S. Coleri, and P. Varraiya. QoS aware adaptive resource allocation techniques for fair scheduling in OFDMA based broadband wireless access systems. *IEEE Transactions on Broadcasting*, December 2003.

[8] G. Zhang. Subcarrier and bit allocation for real-time services in multiuser OFDM systems. *Proc. of IEEE International Conference on Communications* , June 2004.

[9] D-H. Kim, B-H. Ryu, and C-G. Kang. Packet scheduling algorithm considering a minimum bit rate for non-realtime traffic in an OFDMA/FDD-based mobile internet access system. *ETRI Journal*, 26(1), February 2004.

[10] P. Svedman, S. Wilson, and B. Ottersten. A QoS-aware proportional fair scheduler for opportunistic OFDM. *Proc. of Vehicular Technology Conference (VTC)*, September 2004.

[11] J. Gross, H. Karl, F. Fitzek, and A. Wolisz. Comparison of heuristic and optimal subcarrier assignment algorithms. *Proc. of Intl. Conf. on Wireless Networks (ICWN)*, June 2003.

[12] J. Gross, J. Klaue, H. Karl, and A. Wolisz. Subcarrier allocation for variable bit rate video streams in wireless OFDM systems. *Proc. of Vehicular Technology Conference (VTC)*, Florida, October 2003.

[13] Y. Zhang, and K. Letaief. Adaptive resource allocation and scheduling for multiuser packet-based OFDM networks. *Proc. IEEE International Conference on Communications (ICC'04)*, June 2004.

[14] R. Knopp and P. A. Humblet. Information capacity and power control in single-cell multiuser communications. *Proc. IEEE International Conference on Communications (ICC'95)*, Seattle, USA, June 1995.

[15] J. Proakis. *Digital communications*. Mc-Graw Hill Publishing, 1995.

[16] H. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, no(2):83-97, January 1955.

[17] M. Klerer. Introduction to IEEE 802.20: technical and procedural orientation. *IEEE802.20 PD-04*, March 2003.

[18] IEEE C802.20-03/80. Traffic models for IEEE 802.20 MBWA system simulations v.1. September 2003.

[19] IEEE P802.20-04/08 "802.20. SRD: data rates, spectral efficiency and frequency assignments. January 2004.

[20] The IT++ Library, available at http://itpp.sourceforge.net.

[21] Arizona State University Trace Website, Video Library. *Jurassic Park One.*, Available at http://trace.eas.asu.edu/.

[22] S. Shakkottai and A.L. Stolyar. Scheduling algorithms for a mixture of real-time and non-real-time data in HDR. *Proc. of the 17th International Teletraffic Congress - ITC-17*, Salvador da Bahia, Brazil, 793-804, September, 2001.