QoS Aware Dynamic Resource Allocation in Multiuser OFDM Systems with Priority for Different Services

Hamid M. Ramezani
College of Electrical Engineering
Iran Univ. of Science and Technology
Tehran, Iran
hamid.ramezani@gmail.com

Reihaneh Lavafi
College of Electrical Engineering
Iran Univ. of Science and Technology
Tehran, Iran
reihaneh_lavafi@ee.iust.ac.ir

Bahman Abolhassani
College of Electrical Engineering
Iran Univ. of Science and Technology
Tehran, Iran
abolhassani@iust.ac.ir

Abstract—In this paper, we propose an algorithm for dynamic resource allocation in OFDM (orthogonal frequency division multiplexing) or OFDMA (OFD multiple access) based networks. We consider two groups of users consisting of group 1 called constant bit-rate (CBR) users (who are delay-sensitive and data-rate sensitive) and group 2 called best effort (BE) users. The proposed algorithm gives higher priority to the first group at the time of resource (transmitted power levels and subchannels) allocation to each user, to meet their CBR requirements. The performance of the proposed algorithm is evaluated through computer simulations, which show that the algorithm performs near optimum with the added superiority of meeting the bit requirements of the first group users.

I. INTRODUCTION

Nowadays many communication systems, such as WLAN, WMAN and DVB, utilize orthogonal frequency division multiplexing (OFDM) or multiple access (OFDMA) modulation. OFDM accompanied multiple antennas at transceiver sides, is a promising technique for the next generation of wireless communications [1]. Multi-user OFDM provides multiple accesses for OFDM users by allowing them to share an OFDM symbol.

Dynamic resource allocation techniques, which allocate the subchannels and power level to users dynamically, have higher performance on fixed resource allocation schemes, due to the time varying nature of wireless channels. The algorithms are categorized in two main groups: rate adaptive and margin adaptive [5].

Resource allocation for rate adaptive algorithms can be described as maximizing overall throughput, while satisfying requirements of fairness, subject to power restriction and QoS requirements [3, 10]. In the other hand the margin adaptive algorithms try to minimize the overall transmitting power while meeting users’ requirements [5].

Heterogenous services define different kinds of QoS at the systems. Each type of service needs its own data rate and performance (BER and delay). Such requirements add more complexity to the resource allocation algorithms and make them extremely difficult for optimization. Anas et. al., in [6], propose an efficient algorithm for power distribution among two types of OFDMA users, but their subchannel allocation method is inefficient since it is fixed in time. In [7] like [6], K. Kim et. al. try to maximize the total system capacity with the constraints of data-rate sensitive traffics and a newly considered delay sensitive service. Our work is an extension of the work in [6], but we consider further constraints of priority for the services and users, as well as proportional fairness, [8], for the best effort service users.

In this paper, we propose a two step approach to satisfy a higher priority for delay-sensitive and rate-sensitive services while providing both maximum system capacity (rate adaptive algorithm) and proportional fairness among best effort users. We subdivide the hard combinational algorithm into two steps: (1) subchannel allocation based on three factors: the data rate, service priority, and each of subchannel profile; and (2) optimize the power allocation to maximize the total data rate based on users’ constraints.

The remainder of this paper is organized as follows. In Section II, the system model is considered and the objective function and the constraints are formulated. In Section III, the proposed two step algorithm is described. In Section IV, numerical results are presented and we finally conclude the paper in Section V.

II. SYSTEM MODELS

A multiuser OFDMA system is shown in Fig. 1. In the base station, all subchannel profiles are sent to subchannel and power allocation algorithm through feedback channels from all mobile users. We assume that the perfect channel state information (CSI) will be available at the transmitter side. The effects of channel estimation error and the imperfection of CSI at the transmitter are analyzed in [9].

We suppose the total of K users in the system sharing N subchannels. The users have three types of services: delay sensitive, data-rate sensitive and best effort services.
The delay-sensitive users are those who use multimedia applications and as we know the performance of such applications is sensitive to packet delays and data rates. Based on queue theory, we can categorize delay-sensitive users as data-rate sensitive users by mapping their current data rate to a higher one. Like in [7], we model packet arrival as an $M/D/1/d_k$ queue with arrival rate of $\lambda_k$, processing rate $\mu_k$ and buffer size $d_k$. In [7], Li and Lui show that for some $\mu_k > \theta_k$ the delay-sensitivity can be met. For this purpose, they obtain $P_k$, (the probability that a packet delay is more than some given value) and $\theta_k$ (the minimum required data rate for a delay-sensitive service) as given by

$$
P_k = P(\lambda_k, \mu_k, d_k) = \frac{\lambda_k (\mu_k - \lambda_k)}{\lambda_k a^2 + \mu_k s^2},
$$

and

$$
\theta_k = \mu_k^{-1}(\lambda_k, O_k, d_k),
$$

where $a^2$ and $s^2$ are the squared coefficients of variations of the arrival and service process, respectively.

Using the value of $\mu_k$, which is determined by the block, called average rate estimator, we can combine the delay sensitive and data rate users into one group of data-sensitive users. So, in this case, we have two types of users: data-sensitive users and best effort users. Moreover, we assign more priority to the newly combined class of $K_1$ data-sensitive users whose satisfaction is met by providing a constant data rate required by each of them at each time slot. The system tries to meet the data rates of these data-sensitive users in any case, even by sacrificing the resources belong to lower priority group of users served as best-effort. We also rank data-sensitive users from 1 to $K_1$ and give more priority to lower indexed users. This user priority is taken into account at the time of power allocation. We also suppose a constant data rate of $R_k = \mu_k$ for each of these $K_1$ CBR users.

Mathematically, the optimization problem considered in this paper is formulated as (2).

$$
\text{Maximize } \sum_{k=1}^{K_1} \sum_{n=1}^{N} \frac{\rho_{k,n} \lambda_k^2}{N_0 B N \Gamma_k} \log_2 \left(1 + \frac{\rho_{k,n} \lambda_k^2}{N_0 B N \Gamma_k}ight),
$$

subject to the following constraints:

$$
i) \sum_{n=1}^{N} \frac{\rho_{k,n}}{N} \log_2 \left(1 + \frac{\rho_{k,n} \lambda_k^2}{N_0 B N \Gamma_k}ight) = R_k \text{ b/s/Hz for } k \leq K_1
$$

$$
ii) \sum_{k=1}^{K} \sum_{n=1}^{N} \rho_{k,n} \mu_k \leq \rho_{\text{total}}
$$

$$
iii) \rho_{k,n} \geq 0 \text{ for all } k, n
$$

$$
iv) R_{K_1+1} : R_{K_1+2} : \ldots : R_K = \gamma_{K_1+1} : \gamma_{K_1+2} : \ldots : \gamma_K
$$

$$
v) \sum_{k=1}^{K} \rho_{k,n} = 1, \rho_{k,n} \in \{0,1\}, n = 1, \ldots, N
$$

where, $B$ is the total bandwidth; $N$ is the total number of subchannels; $P_{\text{total}}$ is the total available power; $\rho_{k,n}$ is the amount of power allocated to subchannel $n$ of user $k$; $\{\gamma_{k+1}, \ldots, \gamma_K\}$ is a set of values proportional to the best effort users rate. $\Gamma_k$ is the gap SNR required to meet a specific BER for user $k$ [4]. We also consider that each subchannel will be
assigned to only one user in the last constraints and as a result there is no inter-channel interference.

III. PROPOSED SUBOPTIMAL ALGORITHM

Finding the optimal solution of the optimization problem, given by (2), is much complicated, since we should consider both subchannel allocation and power distribution simultaneously. To overcome this problem, we divide the problem into two steps, in step one, the subchannel is allocated assuming equal power for all subchannels and in step two, the optimal solution for power distribution, assuming allocated subchannels, is considered.

A. Subchannel Allocation

Assume that, the total normalized system capacity is C and \( R=R_1+R_2+\ldots+R_K \). We need \( R \) b/s/Hz for the guaranteed service users and the remainder, \( C-R \), for the best effort users, so we can convert the optimization problem in (2) to a problem similar to one given in [8]. The proposed algorithm defines a proportional coefficient \( \gamma_k \) for the guaranteed users like those defined for the best effort users as follows:

\[
\gamma_1 : \gamma_2 : \ldots : \gamma_{K_1} = 1/C \times (R_1 : R_2 : \ldots : R_{K_1})
\]

(3)

Then, it updates the best effort proportional coefficients such that they use the other resources as follows:

\[
\gamma_{K+1}^{new} : \gamma_{K+2}^{new} : \ldots : \gamma_{K}^{new} = \frac{C-R}{C} \times (\gamma_{K+1} : \gamma_{K+2} : \ldots : \gamma_{K})
\]

(4)

By the new defined coefficient for both BE and CBR users, all users in the network have proportional coefficients to zero. Here, one problem is remained: we don’t know the exact value of the total system capacity, \( C \), however we can compute the maximum achievable system capacity, \( C_{max} \), of the system like in [4]. We propose considering \( C=\alpha \times C_{max} \) where \( \alpha \) is a weighing factor less than unity. In the iterative mode, the algorithm can adapt the value of \( \alpha \), and hence \( C \), in each iteration. This helps finding solutions very close to real system capacity much faster. After finding the appropriate proportional coefficients, \( \gamma_k \), of all users, the assignment of subchannels to all users, explained below, is initiated.

1) Initialization
   a) Set \( R_k=0, \Omega_k=\Phi \) for \( k=1 \) to \( K \) and \( A=\{1, 2, \ldots, N\} \)
   b) for \( k=1 \) to \( K \)

2) While \( A \neq \Phi \)
   a) Find \( n \) satisfying \( |H_{k,n}| \geq |H_{k,j}| \) for all \( j \in A \).
   b) Let \( \Omega_k=\Omega_k \cup \{n\}, A=A-\{n\} \) and update \( R_k \).

   3) While \( A \neq \Phi \)
      a) Find \( k \) satisfying \( R_k/\gamma_k \leq R_i/\gamma_i \) for all \( i, 1 \leq i \leq K \).
      b) For the found \( k \), find \( n \) satisfying \( |H_{k,n}| \geq |H_{k,j}| \) for all \( j \in A \).
      c) For the found \( k \) and \( n \), Let \( \Omega_k=\Omega_k \cup \{n\}, A=A-\{n\} \) and update \( R_k \) where \( H_{k,n} \) is defined as \( (p_k h^k_n)/(N_0 \beta_k \Gamma_k/N) \).

By such allocation we jointly distribute the subchannel among users with different services. So the subchannels among CBR users and BE are allocated dynamically and efficiently, and both parts can utilize any subchannel independent of their services. Relative to partitioning subchannels for different services [6], the proposed scheme allocates the resources more efficiently.

B. Optimal Power distribution for fixed subchannel Allocation

For a certain allocated users’ subchannels, which is done in subchannel Allocation section, the optimization problem in (2) is formulated as:

\[
\max \sum_{k=1}^{K} \sum_{n=\Omega_k} \frac{1}{N} \log_2 \left( 1+ \frac{p_k,n}{N_0 B/N \Gamma_k} \right)
\]

(5)

subject to the following constraints:

i) \( \sum_{n=\Omega_k} \frac{1}{N} \log_2 \left( 1+ \frac{p_k,n h^2_{k,n}}{N_0 B/N \Gamma_k} \right) = R_k \) b/s/Hz for \( k \leq K^1 \)

ii) \( \sum_{k=1}^{K} \sum_{n=\Omega_k} p_{k,n} \leq P_{total}, \quad p_{k,n} \geq 0 \) for all \( k, n \) and

iii) \( R_{K+1} : R_{K+2} : \ldots : R_K = \gamma_{K+1} : \gamma_{K+2} : \ldots : \gamma_K \),

where \( \Omega_k \) is a set of indices of subchannels allocated to user \( k \).

By the help of Lagrangian optimization, the problem in (5) is equivalent to finding the maximum of the following cost function.

\[
L = \sum_{k=K+1}^{K} \sum_{n=\Omega_k} \frac{1}{N} \log_2 \left( 1+ \frac{p_k,n H_{k,n}}{} \right)
\]

\[
+\alpha_k \left( \frac{P_{total}}{K} \sum_{k=1}^{K} \sum_{n=\Omega_k} p_{k,n} + \frac{K_1}{K} \sum_{k=1}^{K} \alpha_k \left( \frac{1}{N} \log_2 \left( 1+ \frac{p_k,n H_{k,n}}{} \right) - \gamma_k \right) \right)
\]

+ \( \sum_{k=K+1}^{K} \alpha_k \left( \sum_{n=\Omega_k} \frac{1}{N} \log_2 \left( 1+ \frac{p_{K+1,n} H_{K+1,n}}{} \right) - \gamma_k \right) \)

(6)
where \( \alpha_k, k=0, 2, \ldots, K+1 \), are the Lagrangian coefficients.

By differentiating (6) with respect to \( p_{k,n} \) and setting each derivative to zero for constant data-rate users, we obtain the power allocated to \( k \)th CBR user, in subchannel \( n \) as given by (7). We assume user \( k \) utilize \( N_k \) subchannel and its subcarrier are arranged according to the ascending orders of carrier to noise ration (CNR) i.e. \( H_{k,1}<H_{k,2}<\ldots<H_{k,N_k} \)

\[
P_{k,n} = \left( \frac{2^{R_k} N_k \prod_{m \in \Omega_k} H_{k,n}^{1/N_k} - 1}{H_{k,n}} \right), \text{ for } k \leq K+1
\]

where \( N_k \) is the number of subchannels allocated to user \( k \).

Using (8), for best-effort users, we obtain

\[
P_{k,n} = p_{k,K+1} + \frac{H_{k,n}-H_{k+1,n}}{H_{k,n} H_{k+1,n}},
\]

with the following constraint

\[
\frac{1}{N_{K+1} N} \sum_{n=1}^{N_{K+1}} \left( \log_2 \left( 1 + \frac{P_{k+1,n} H_{k,n}}{N_k} \right) + \log_2 \left( W_{k+1,n} \right) \right)
\]

\[
= \frac{1}{N_k N} \sum_{n=1}^{N_k} \left( \log_2 \left( 1 + \frac{P_k H_{k,n}}{N_k} \right) + \log_2 \left( W_k \right) \right) \text{ for } k > K+1
\]

where in equation (9) \( P_k, W_k, \) and \( V_k \), are defined by (10) to (12).

\[
P_k = \sum_{k=K+1, K+2, \ldots}^{N_k} p_{k,n},
\]

\[
V_k = \sum_{n=2}^{N_k} H_{k,n} H_{k,n,1}
\]

\[
W_k = \frac{1}{\prod_{n=2}^{N_k} H_{k,n,1}^{1/N_k}}
\]

In the above formula, it is assumed that \( p_{k,n} \) is greater than zero. If \( p_{k,n} \) is negative, the subchannel allocation must be repeated in a more efficient procedure.

In power distribution among CBR users, the power allocated to each of users’ subchannels is the minimum required power to meet their data rate. The priority of CBR users is higher than the BE users. At first, the power is dedicated to the CBR users, and the remaining power will be distributed to BE users through (8) to (12). If there is no extra power, the data rate for BE users will be zero (in this case their proportional coefficients are zeros). As was mentioned in Section 2, a higher priority is given to lower indexed users who are ranked from 1 to \( K_1 \). The process of power allocation is continued until the total power of the base station transmitter is exhausted. So, at some instants, some lower priority data-rate users gain no power at all.

It can be seen that, the better the subchannel transmission gain is, the more power is dedicated to it. This is a famous water-filling algorithm used for group of BE user’s subchannels [11].

In the next section we evaluated the efficiency of the proposed algorithm in satisfying the bit-rate, and priority of different kinds of users. The total system throughput is much of attention too.

IV. NUMERICAL RESULTS

In this section, through computer simulations we evaluate the average data rate normalized by the total bandwidth. We assume that each user’s subcarrier signal undergoes independent Rayleigh fading channels whose statistics are identical such that the average power gain, \( E[|\alpha_k|^2] \), of each subchannel equals to unity. The average SNR (ASNR) is defined as \( P_{total}/(N_0 B) \) with the fixed total bandwidth \( B \). The required bit error rate (BER) of all users is set to \( 10^{-3} \) (\( \Gamma=1.87 \)). The number of subchannels, \( N \), is chosen as 128 and the weighting factor, \( \omega \), is set to be 1. To obtain the average data rate, we have simulated 1000 independent trials, and for the iterative part (updating the total system capacity and consequently the users modified proportional factor) maximum of 10 iterations is chosen.

In Fig. 2 the performance of the proposed algorithm is compared with the maximum achievable system capacity in OFDMA system [4], and a fixed resource allocation scheme named TDMA-Water filling [2]. In [4] the given subchannel will be allocated to a user who has a higher carrier to noise ratio at that subchannel so the probability of fading in each subchannel goes to lowest level and consequently the total system capacity reaches to the highest possible one. The proof
is studied in [4]. TDMA-Water filling is a kind of fixed resource allocation which repeatedly allocates all the subchannels to a user at dedicated time slots. In TDMA-WF (water filling) at each time slot only one user is served by all the subchannels.

There are $K_1 = 5$ CBR users along with 5 BE users. The bit rate required for each of the $K_1$ users are shown in Fig. 2. As well, the total constant bit rate required for the $K_1$ users is considered to be smaller than the total system capacity of the system. As it is seen in this figure, the normalized total system capacities of both the proposed non-iterative and iterative algorithms are fallen on each other and they are near to $C_{\text{max}}$. As the signal to noise ratio of the system increases, the normalized total capacity of the system increases too. It must be mentioned that the users’ constraints are satisfied by the proposed algorithms.

To better understand the operation of the proposed algorithm, the normalized user capacity of each of the two types of users is also shown in Fig. 3. As mentioned before, the algorithm first allocates the bit rate to each of the CBR users, and then, if any resources (power and subchannels) left unused are proportionally allocated to the BE users. For this reason, by increasing ASNR, the capacities of CBR users remain constant while the capacity of each of BE users increases.

There are many other scenarios that can be considered for evaluation of the proposed algorithms. For example, other different constant bit rates may be considered for CBR users.

In Fig. 4, we investigate the effect of increasing the number of CBR users while sum of their normalized capacity is unchanged and is equal to one. As the number of CBR users increases the effect of multiuser diversity of the OFDMA system causes the total maximum achievable capacity of the system increases, but it has no effect on the proposed algorithm and the total system capacity increases a little. It is seen that the iterative approach performs a little better.

In Fig. 5, effects of increasing the number of BE users are studied. It is assumed that ten CBR users are in the system and their total constant bit rate is unchanged during simulation. The average SNR is 12 dB. From this figure we can understand that, the more the number of the BE users the more multiuser diversity gain is achieved and as a result, the total system capacity of the proposed algorithm increases and gets closer to the $C_{\text{max}}$.

In Fig. 6, the effect of increasing total constant bit rate is investigated. As it is shown in this figure, there is a region (with 1 to 1.5 bits/sec/Hz) that the algorithm performs near optimum. It is also shown that the system performance degrades as the total constant capacity becomes larger than a specific amount. The algorithm ignores lower priority users among CBR users when it senses that, it is unable to provide needed power for them. But in such a case some subchannels become unused, because they were allocated to these lower priority users in step 1. So the subchannel allocation must be recomputed considering such event.

As seen in Fig. 6, the iterative approach performs better than non-iterative algorithm when the resources are low to meet the constraints. From the figure, it is concluded that we
shouldn’t let the system work in the degraded performance region by managing the CBR users’ total capacity in the near optimum region. In Fig. 7, the capacity of each of users in this case is shown. Ignoring lower priority users is obvious in the figure.

Figure 6. Effect of increasing the total constant bit rate

![Figure 6](image)

Figure 7. Capacity of users versus total constant capacity

![Figure 7](image)

V. CONCLUSION

In this paper, we considered heterogeneous services: constant-bit-rate (delay and data-rate sensitive) and best-effort users. We proposed an efficient algorithm to allocate subchannels and power to all users in the system. The performance of the proposed algorithm was evaluated by computer simulations, which show that the system can perform well by meeting the requirement of delay and data-rate sensitive users. Moreover, managing the total capacity of CBR users helps system work near maximum achievable capacity, while the requirements of all CBR users are also met. It is also shown that increasing the number of BE users (relative to the CBR users) results in near to optimum performance for the proposed algorithm with the added superiority of meeting the CBR users’ requirements.

ACKNOWLEDGMENT

The authors would like to express their sincere thanks to the Iran Telecommunication Research Center (ITRC) for supporting this research.

REFERENCES