Abstract—In this paper, resource allocation for heterogeneous services is studied in multiuser orthogonal frequency division multiplexing (OFDM) systems. We propose a resource allocation algorithm, which is designed to improve the system throughput while satisfying the quality of service (QoS) requirements of both the real-time and nonreal-time services. In the proposed algorithm, the resources, composed of subcarriers and transmit power, are adaptively allocated to the users based on their service types and channel states over two sequential steps: (1) resource allocation for the real-time users that minimizes the resource usage required to satisfy the data rate requirement, and (2) resource allocation for the nonreal-time users that maximizes the system throughput using the remaining resource. The performance of the proposed resource allocation algorithm is evaluated in a frequency-selective fading channel, and compared with that of a simple resource allocation algorithm. Numerical results show that the proposed algorithm provides a significant throughput gain over the conventional algorithm.

I. INTRODUCTION

As demand for broadband mobile communication increases, mobile communication systems are expected to support various types of services simultaneously. Each type of service has its own quality of service (QoS) requirements such as delay, data rate, and bit error rate, etc. A resource allocation algorithm is needed to utilize resources efficiently while satisfying the QoS requirements of each service. In this work, we investigate a resource allocation algorithm suitable for supporting the real-time and nonreal-time services simultaneously.

Based on the delay requirement, we can broadly classify communication services into two categories: real-time and nonreal-time services. Since real-time services such as voice have a strict delay requirement, a resource allocation should guarantee the constant data rate [1]. On the contrary, nonreal-time services such as file transfer allow relatively large delay. In this case, it is desirable that a resource allocation algorithm maximizes the system throughput [2]-[5]. Although those above-mentioned schemes have dealt with the problem of resource allocation, they consider the situation of supporting only a single type of service. To accommodate various types of service simultaneously under limited resources, we are motivated to investigate a different resource allocation strategy.

In this paper, we investigate resource allocation for a multiuser orthogonal frequency division multiplexing (OFDM) system that supports both the real-time and nonreal-time services together. We propose a subcarrier and power allocation algorithm to maximize the system throughput while satisfying QoS requirements of both real-time and nonreal-time services. The proposed resource allocation algorithm operates over two steps: (1) resource allocation for the real-time services and (2) resource allocation for the nonreal-time services. In order to maximize the system throughput, the proposed resource allocation algorithm allocates resource to real-time users such that the allocated resource is minimized in the first step. In the second step, the remaining resource is allocated to nonreal-time users in a way that maximizes the throughput.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A configuration of multiuser OFDM system that supports both real-time and nonreal-time services is depicted in Figure 1. The system serves $K_R$ real-time users and $K_{NR}$ nonreal-time users, and thus the total number of users in the system is $K = K_R + K_{NR}$. At the transmitter, the data streams from $K$ users are fed into the resource allocation block, which allocates subcarriers and transmit power to the users. The channel state information is assumed to be delivered from each user to the transmitter through an error-free feedback channel. Based on the channel state information, the resource allocation block allocates $N$ OFDM subcarriers and transmit power $P$ to $K$ users. Then, data streams on each subcarrier are passed to the adaptive modulator, which determines a modulation level. The modulated signals are transmitted after inverse fast Fourier transform (IFFT), parallel to serial (P/S) conversion, and guard interval insertion. We assume that the resource allocation information is passed to the receiver of users through an appropriate control channel, so that the data stream for each user can be recovered from the received signal at the receiver.

The goal of resource allocation is to allocate the subcarriers and transmit power so that the system throughput is maximized, while satisfying the QoS requirements of each user. We consider two QoS requirements: data rate and bit error rate (BER) requirements. Data rate requirement of $R_k$ bits per OFDM symbol duration and the BER requirement of $BER_k$ are imposed upon the $k$th real-time user. For nonreal-time users, we assume that there is no data rate requirement but there is the BER requirement of $BER_k$ for the $k$th nonreal-time user. Regardless of real-time or nonreal-time service, when the $n$th subcarrier is used by the $k$th user, the achievable data rate may be expressed as [5]

$$c_{k,n} = \log_2 \left( 1 + \frac{P_{k,n} g_{k,n}}{\sigma^2} \right) \text{ bits/OFDM symbol time} \quad (1)$$

where $P_{k,n}$ and $g_{k,n}$ are, respectively, the allocated power and channel gain for the $k$th user and the $n$th subcarrier, and $\sigma^2$ denotes the noise power. $\Gamma_k$ is the required signal to noise ratio (SNR) for achieving $BER_k$, and may be written as [5]

$$\Gamma_k = -\ln(5BER_k) \cdot \frac{1.5}{1.5} \quad (2)$$

This work was supported by the National Research Laboratory Program of Korea.
Let us introduce $\rho_{k,a}$, referred to as the subcarrier allocation indicator, to show whether the $n$th subcarrier is allocated to the $k$th user or not:

$$\rho_{k,a} = \begin{cases} 0, & \text{if } c_{k,a} = 0 \\ 1, & \text{otherwise.} \end{cases}$$

(3)

As in [5], we assume that each subcarrier can be allocated to only one user. This implies that $\rho_{k,a}$ satisfies $\sum_{k=1}^{K} \rho_{k,a} = 1$ for all $a$. In this case, the achievable data rate of the $k$th user may be calculated as

$$D_k = \sum_{a=1}^{N} \gamma_{k,a} \cdot c_{k,a} = \sum_{a=1}^{N} \rho_{k,a} \cdot \log_2 \left( 1 + \frac{P_{k,a} \cdot g_{k,a}}{\sigma^2} \right),$$

(4)

and the system throughput is given as

$$T = \sum_{k=1}^{K} D_k.$$  

(5)

In summary, the resource allocation considered in this paper can be formulated as

$$\arg \max_{\{\rho_{k,a}\}} \left[ T \right] = \arg \max_{\{\rho_{k,a}\}} \left[ \sum_{k=1}^{K} \sum_{a=1}^{N} \rho_{k,a} \cdot \log_2 \left( 1 + \frac{P_{k,a} \cdot g_{k,a}}{\sigma^2} \right) \right],$$

(6)

where the maximization is subject to the constraints of

(C1) total power constraint: $\sum_{k=1}^{K} \sum_{a=1}^{N} P_{k,a} = P$

(C2) total number of subcarriers constraint: $\sum_{k=1}^{K} \sum_{a=1}^{N} \rho_{k,a} = N$

(C3) real-time users’ data rate requirement: $D_k = R_k$ if the $k$th user is a real-time user.

(C4) BER requirement: $BER_k, k = 1, 2, \cdots, K$.

### III. RESOURCE ALLOCATION ALGORITHM

In this section, we propose a resource allocation algorithm for real-time and nonreal-time users. For notational simplicity, we assume that the first to the $K$th users are real-time users: $k = 1, 2, \cdots, K$. Automatically, the $(K+1)$th to the $K$th users are the nonreal-time users. Then, the system throughput in (5) can be rewritten as

$$T = \sum_{k=1}^{K} D_k + \sum_{k=K+1}^{K} D_k,$$

(7)

where the first term denotes the throughput of real-time users, and the second one is that of nonreal-time users. Note that the first term is constant due to the data rate requirements of real-time users, $D_k = R_k$, $k = 1, 2, \cdots, K$. Thus, maximizing $T$ in (7) is equivalent to maximizing the second term. In order to maximize the second term, we need to maximize the amount of resource for nonreal-time users, which can be achieved by minimizing the resource usage of real-time users satisfying $D_k = R_k$. Based on this observation, we propose a two-step resource allocation algorithm. In the first step, the proposed algorithm minimizes the resource usage of real-time users under the constraint of $D_k = R_k$. In the second step, the remaining resource is allocated to nonreal-time users so that the overall throughput is maximized. The proposed algorithm is described in more detail in the following two subsections.

#### A. The First Step: Resource Allocation for Real-time Users

In this subsection, we consider the minimization of resource usage needed for achieving real-time users’ data rate requirements. Assume that there exist only real-time users. Note that there is a tradeoff relationship between power and subcarrier usages; if we decreases the allocated power for a user, then we need to allocate more subcarriers to achieve the required data rate, and vice versa. Hence, the resource usage should be counted in terms of both the power and subcarrier usages. Accordingly, we define the resource usage of the $k$th real-time user as

$$\eta_{k} = \frac{\sum_{a=1}^{N} P_{k,a} \cdot s_{k,a}}{P}, \quad k = 1, 2, \cdots, K$$

(8)

where the first term denotes the normalized power usage, and the second one is the normalized subcarrier usage. Then, the minimization of resource usage for real-time users can be expressed as

$$\arg \min_{\{s_{k,a}\}} \left[ \sum_{k=1}^{K} \eta_{k} \right] = \arg \min_{\{s_{k,a}\}} \left[ \sum_{k=1}^{K} \left( \frac{\sum_{a=1}^{N} P_{k,a} \cdot s_{k,a}}{P} + \frac{\sum_{a=1}^{N} P_{k,a} \cdot s_{k,a}}{N} \right) \right],$$

(9)

subjected to the data rate requirements $D_k = R_k$, $k = 1, 2, \cdots, K$. This minimization should be done by jointly considering all users, which makes the problem difficult to solve. As a suboptimal approach, we propose an iterative algorithm. At each iteration, the proposed algorithm selects one user and then calculates the number of subcarriers and the amount of power that minimize the resource usage of the selected user. If a user’s channel condition is better than the others’, it is desirable to select that user since less power and fewer subcarriers are needed to achieve a certain data rate. We use the number of best subcarriers as the representative of channel condition of each user. The number of best subcarriers for each user is defined as the number of subcarriers, of which channel gains for that user are larger than those for the others. We define the set of best subcarriers for the $k$th user as

$$s_k = \{ n : \arg \max_{\{s_{k,a}\}} \{ g_{k,a,n} \} = \{ k \} \}.$$  

(10)

Then, the selected user at the first iteration corresponds to the user with the largest cardinality $|s_k|$ of $s_k$, and the index of the selected user is given as

$$k^* = \arg \max_{k} \{|s_k| \}, \quad k = 1, 2, \cdots, K.$$  

(11)

After selecting one user, we determine the number of subcarriers and the amount of power that minimize the resource usage of the $k^*$ th user in (8). We use the exhaustive search method to find the required number of subcarriers. Note that the maximum available number of subcarriers for the $k^*$ th user is $|s_{k^*}|$, and the corresponding channel gains are $g_{k^*,n}, n \in s_{k^*}$. Suppose that the requirements of the $k^*$ th user can be satisfied using $m$ subcarriers among $s_{k^*}$ subcarriers. Then, it is desirable to use $m$ subcarriers that correspond to the $m$ largest channel gains. With these $m$ subcarriers, the resource usage of $k^*$ th user is found as

$$\eta_{k^*,m} = \frac{\sum_{i=1}^{m} P_{k^*,i} \cdot s_{k^*,i}}{P}, \quad m = 1, 2, \cdots, |s_{k^*}|.$$  

(12)

where the subscript $n(i)$ denotes the subcarrier index for the $i$th largest channel gain among $\{g_{k^*,n} \}, n \in s_{k^*}$. Obviously, to minimize $\eta_{k^*,m}$ in (12), the sum of allocated power $\sum_{i=1}^{m} P_{k^*,i}$ should be minimized, and the allocated power can be found using the Lagrange multiplier method with the Lagrangian...
where \( \lambda_n \) is the Lagrange multiplier. It is well known that the minimum allocated power can be obtained from the water-filling solution:

\[
P_{(m)}^{\ast,\ast(i)}(i) = \left( \ln 2 \sigma^{-2} \Gamma_{k} \right) \frac{g_{k\ast}^{\ast,\ast(i)}}{\sigma^2 \Gamma_{k}}, \quad i = 1, 2, \ldots, m, \tag{14}
\]

where \((x)\) denotes \( \max(x, 0) \). With these \( P_{(m)}^{\ast,\ast(i)}(i) \), the resource usage \( \eta_{\ast\ast, it} \) in (12) may be calculated for a given \( m \). Finally, the minimum resource usage may be obtained by choosing the minimum \( \eta_{\ast\ast, it} \):

\[
\eta_{\ast\ast, \min} = \min \{ \eta_{\ast\ast, it}, m = 1, 2, \ldots, |s_t| \}. \tag{15}
\]

Accordingly, the number of subcarriers that minimizes the resource usage for the \( k \)th user may be expressed as

\[
m' = \arg \min_{m} \{ \eta_{\ast\ast, it}, m = 1, 2, \ldots, |s_t| \}. \tag{16}
\]

The allocated subcarrier indices are \( n(1), n(2), \ldots, n(m') \), and the allocated power can be obtained by (14) with replacing \( m \) with \( m' \).

For the following \( \mathcal{K}_{R} - 1 \) iterations, we apply the similar procedures of the first iteration; select one user and then calculate the number of subcarriers and power to minimize the resource usage in (12). When selecting one user, the users and subcarriers allocated in the previous iterations should be excluded. In a particular case, there can be real-time users whose data rate requirements are not satisfied. In this case, we drop these users and move to the next iteration.

### B. The Second Step: Resource Allocation for Nonreal-time Users

In this subsection, we consider resource allocation for nonreal-time users that allocates the remaining resources for the second term in (7) to be maximized. The power and subcarriers allocated to the real-time users are given as

\[
P_{R} = \sum_{k=1}^{\mathcal{K}_{R}} \sum_{n=1}^{\mathcal{N}_{R}} P_{k,n}, \tag{17}
\]

\[
N_{R} = \sum_{k=1}^{\mathcal{K}_{R}} \sum_{n=1}^{\mathcal{N}_{R}} P_{k,n}. \tag{18}
\]

Then, the remaining power and subcarriers for the nonreal-time users are given as

\[
P_{\mathcal{R}} = P - P_{R}, \tag{19}
\]

\[
N_{\mathcal{R}} = N - N_{R}. \tag{20}
\]

To maximize the system throughput, the subcarrier should be allocated to only one user who has the best channel gain for that subcarrier [5]. Hence, we allocate each remaining subcarrier to one nonreal-time server whose channel gain is the largest among \( K_{\mathcal{R}} \) channel gains. From the allocated users’ channel gain, we can easily allocate the power by using the Lagrange multiplier method similar to that in [5].

Let the set of \( N_{\mathcal{R}} \) remaining subcarriers be \( \mathcal{S} \). Then the allocated power becomes

\[
P_{k,n, \mathcal{S}} = \begin{cases} \frac{\lambda}{\ln 2} \frac{\sigma^2 \Gamma_{k}(s)}{\sum_{k=1}^{(\mathcal{K}_{R}+1), \mathcal{S}(k), \mathcal{S}(n)} \max_{k,n}} \quad & \text{if } n \in \mathcal{S} \vphantom{\frac{\Gamma_{k}(s)}{\sum_{k=1}^{(\mathcal{K}_{R}+1), \mathcal{S}(k), \mathcal{S}(n)}}} \\ 0, \quad & \text{otherwise} \end{cases} \tag{21}
\]

where the subscript \( k(n) = \arg \max \left\{ g_{k,n}, \ k = \mathcal{K}_{R}+1, \mathcal{K}_{R}+2, \ldots, \mathcal{K} \right\} \)

The Lagrange multiplier \( \lambda \) is determined by the sum power constraint \( \sum_{k=1}^{\mathcal{K}_{R}} P_{k,n, \mathcal{S}} = P_{\mathcal{R}} \).

### IV. Numerical Results

In this section, we evaluate the performance of the proposed resource allocation algorithm described in Section III. Assume that all real-time users have the same BER requirement and data rate requirements of \( BER_k = 10^{-3} \) and \( R_k = 5 \), respectively. Similarly, all nonreal-time users have the same required BER of \( 10^{-2} \): \( BER_k = 10^{-3} \) for \( k = \mathcal{K}_{R}+1, \mathcal{K}_{R}+2, \ldots, \mathcal{K} \). The channel gains \( g_{k,n} \) are assumed to be Rayleigh distributed and independent of each other. In the following, we define the total SNR as \( P/\sigma^2 \) and evaluate the average system throughput through 100,000 independent simulation runs. The performance of the proposed resource allocation algorithm is compared with that of a simple resource allocation algorithm. Since there is no previous resource allocation algorithm for heterogeneous services, we introduce a simple resource allocation algorithm referred to as the fixed number of subcarriers allocation. Assuming that each real-time user is allocated with \( M \) subcarriers, then this algorithm allocates \( M \) subcarriers with the \( M \) largest channel gains for \( \mathcal{K}_{R} \) real-time users. The subcarriers already allocated to some other users will be excluded from the selection of \( M \) subcarriers for the following users. After selecting \( M \) subcarriers, the allocated power may be obtained by the water-filling method similar to (14). Then the remaining subcarriers of \( N - M \) power are allocated to the nonreal-time users by following the same method as in Section III-B. For simplicity, we assume \( M = 1 \) throughout this section.

Figure 2 shows the average system throughput of the proposed resource allocation algorithm for various total SNR, and compare it with that of the fixed allocation algorithm, when \( \mathcal{K}_{R} = 8, \mathcal{K}_{\mathcal{R}} = 8, \) and \( N = 64 \). The proposed algorithm is shown to provide higher average system throughput than the fixed allocation algorithm. At low SNR, the average system throughputs become almost the same for the two resource allocation algorithms, and converge to zero since the available power is small. At the intermediate SNR, the proposed algorithm performs apparently better than the fixed one. As the available power becomes larger, the throughputs of two algorithms are increasing and become the same. The reason is that, for high SNR, it is desired to allocate as small number of subcarriers as possible to the real-time users, since there is a plenty of power to use. Hence, the performance of proposed algorithm becomes the same with that of the fixed number of subcarriers allocation that allocate only one subcarrier to the real-time users.

Figure 3 shows the effect of number of subcarriers on the average system throughput. As the number of subcarriers increases, the average system throughput of the proposed algorithm increases rapidly while that of the fixed allocation does so slightly. This indicates that the proposed algorithm utilizes the resource more efficiently than the fixed allocation.

The effect of number of real-time users on the system throughput is shown in Figures 4. Regardless of the number of real-time users, the average system throughput of the proposed algorithm remains constant with slight fluctuation.
However, that of the fixed allocation algorithm rapidly decreases and remains constant for $K_R = 6$. Figure 5 shows the effect of number of nonreal-time users on the system throughput. As the number of nonreal-time users increases, the average system throughput of the proposed algorithm smoothly increases while that of the fixed allocation algorithm remains constant.

V. CONCLUSIONS

In this paper, we investigated resource allocation for heterogeneous services in the multiuser OFDM system. We proposed a resource allocation algorithm, which is designed to maximize the system throughput while satisfying all users’ QoS requirements. The proposed algorithm allocates resource for real-time users first and then for non-real-time users. Regarding real-time users, the proposed algorithm is designed to minimize the required resource usage while satisfying their data rate requirements. The remaining resources are allocated for non-real-time users with the objective of maximizing the throughput. For performance comparison, we introduced a simple resource allocation algorithm referred to as the fixed number of subcarriers allocation. The numerical results showed that the proposed allocation algorithm significantly outperforms the fixed allocation algorithm.

REFERENCES