

Research Article

Dynamic Subcarrier Allocation for Real-Time Traffic over Multiuser OFDM Systems

Fanglei Sun,¹ Mingli You,¹ and Victor O. K. Li²

¹Research and Innovation Center, Alcatel-Lucent Shanghai Bell Co., Ltd, Shanghai 201206, China

²Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong

Correspondence should be addressed to Fanglei Sun, fanglei.a.sun@alcatel-sbell.com.cn

Received 24 January 2009; Accepted 14 April 2009

Recommended by Dmitri Moltchanov

A dynamic resource allocation algorithm to satisfy the packet delay requirements for real-time services, while maximizing the system capacity in multiuser orthogonal frequency division multiplexing (OFDM) systems is introduced. Our proposed cross-layer algorithm, called Dynamic Subcarrier Allocation algorithm for Real-time Traffic (DSA-RT), consists of two interactive components. In the medium access control (MAC) layer, the users' expected transmission rates in terms of the number of subcarriers per symbol and their corresponding transmission priorities are evaluated. With the above MAC-layer information and the detected subcarriers' channel gains, in the physical (PHY) layer, a modified Kuhn-Munkres algorithm is developed to minimize the system power for a certain subcarrier allocation, then a PHY-layer resource allocation scheme is proposed to optimally allocate the subcarriers under the system signal-to-noise ratio (SNR) and power constraints. In a system where the number of mobile users changes dynamically, our developed MAC-layer access control and removal schemes can guarantee the quality of service (QoS) of the existing users in the system and fully utilize the bandwidth resource. The numerical results show that DSA-RT significantly improves the system performance in terms of the bandwidth efficiency and delay performance for real-time services.

Copyright © 2009 Fanglei Sun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Demands for real-time multimedia applications are increasing rapidly for broadband wireless networks. Orthogonal frequency division multiplexing (OFDM) is considered a promising technique in such systems. In this paper, we consider multiuser systems [1] where multiple users are allowed to transmit simultaneously on different subcarriers per OFDM symbol. Mobile users on certain OFDM subchannels may experience deep frequency-selective fading in a multipath propagation environment. Since each user may have a different subchannel impulse response, a poor subchannel for one user may be a good subchannel for another user. Clearly, if a user who suffers from poor subchannel gain can be reassigned to a better subchannel, the total throughput can be increased. This is also known as multiuser diversity. Since the subcarrier gains vary from user to user, to achieve higher system capacity and spectral efficiency, it is better to allocate subcarriers and the corresponding power dynamically according to the instantaneous channel states of all users.

To support QoS for multiple services, packet scheduling has been identified as an important mechanism in wired networks. When considering the multipath fading, high error rate, and time-varying channel capacity in wireless links, some new packet scheduling algorithms are developed, such as channel state dependent round Robin (CSD-RR) [2], feasible earliest due date (FEDD) [3], modified largest weighted delay first (M-LWDF) [4], and link-adaptive LWDF [5] algorithms. CSD-RR schedules the packets whose channel is in the "Good" state in a Round Robin fashion. FEDD focuses on scheduling the packet which has the smallest time to expiry and whose channel is in the "Good" state. M-LWDF schedules the packet according to $\max\{\gamma_j r_j(t) W_j(t)\}$, where $W_j(t)$ is the head-of-the-line packet delay for queue j , $r_j(t)$ is the channel capacity with respect to flow j , and γ_j are arbitrary positive constants. M-LWDF is proven to be a throughput-optimal scheduling algorithm. Link-adaptive LWDF aims to satisfy the stringent packet delay constraints, but without any guarantees. The objectives of these algorithms are to maximize the system spectral efficiency by exploiting the random channel variations and

to provide QoS guarantees to the users by deferring the transmissions on the deep fading links and compensating for them when the links recover. However, all these scheduling algorithms are based on packet scheduling, and multiple frequency subcarrier scheduling, which may be implemented in multiuser OFDM systems, is not considered. In the PHY layer, the total power resource is limited. Given the required number of subcarriers of different users, how to minimize the power allocation for the users on the subcarriers under users' SNR requirements is still a problem. To solve this problem, a suboptimal subcarrier allocation algorithm based on constructive assignment and iterative improvement is proposed in [6] and adopted in [7]. The algorithm exploits the similarity between the subcarrier allocation problem and the classical assignment problem. However, the algorithm can only provide a suboptimal allocation. An optimal solution to this power minimization problem is the Kuhn-Munkres algorithm proposed for the classical assignment problem [8]. Kuhn-Munkres is based on the Hungarian algorithm [9]. OFDM subcarrier allocation using this method has been studied in [10]. However, an important assumption in that paper is that the number of assigned subcarriers for the users is known. Actually, without this information, the Kuhn-Munkres algorithm cannot perform the subcarrier allocation. In addition, in most of the proposed scheduling algorithms, the dynamic variation of the number of active users in the system is ignored.

In this paper, we propose a cross-layer resource allocation scheduling algorithm, named DSA-RT, for real-time services under frequency-selective fading channel in multiuser OFDM systems. This algorithm has two cooperative components: the MAC-layer scheduling/control scheme and the PHY-layer resource allocation scheme. At the MAC layer, based on queuing theory, active users' expected resource requirements to satisfy delay constraints are calculated in terms of the number of subcarriers per OFDM symbol. With the support of our MAC-layer scheduling scheme, the number of required subcarriers and the users' transmission priorities are given. At the PHY layer, based on the modified Kuhn-Munkres algorithm, a PHY-layer resource allocation algorithm is proposed to satisfy all users' requirements under the system SNR and power constraints and to decide the real subcarrier allocation for each active user. (Users admitted to the system are termed active users. Once new users are admitted, they will be allocated resources (subcarriers) by the access control scheme.) When considering a system where the number of active users changes dynamically, if there are still subcarriers left in an OFDM symbol, the access of new mobile users will be considered. In addition, if the dropping rates of certain users violate their maximum tolerable limits, a removal scheme is triggered to remove the aggressive users so as to guarantee the QoS of the other existing users. With the cooperation of the MAC and PHY layer schemes, our proposed algorithm offers the following advantages: (1) based on queuing theory, real-time users' delay requirements can be evaluated in terms of the number of subcarriers required, leading to a more flexible scheduling algorithm which can effectively guarantee the QoS for real-time services in multiuser OFDM systems; (2) with

the number of the expected subcarriers and transmission priority information from the MAC layer, the proposed PHY-layer resource allocation scheme aims to maximize the bandwidth usage under the current channel state, system SNR, and power constraints; (3) when the number of mobile users is dynamically changed, the access control and removal schemes can dynamically adjust system flows and provide delay-related guarantee for the active users in the system.

The rest of this paper is organized as follows. The system model is introduced in Section 2. The detailed description of DSA-RT is presented in Section 3. The simulation results are given in Section 4. Section 5 draws the conclusions.

2. System Model

Figure 1 shows our downlink OFDM system model at a base station (BS). As in previous work [2–5], channel state information (CSI) is assumed to be available at BSs. Assume that the frequency bandwidth is divided into N subcarriers, and there are K active users, where K is changed dynamically and follows a Poisson distribution. BSs are in charge of subcarrier scheduling and resource allocation. We assume a fixed modulation for all subcarriers. The total transmission power is constrained at P and will be optimally allocated to each subcarrier.

BS establishes a queue for each user. Packets are assumed to have equal length of L bits each. Head of line (HOL) packets of queues are scheduled on different subcarriers in different OFDM symbols based on transmission priorities obtained in Section 3. The transmission process for each user can be modelled as an M/G/1 queue. Define the average system time of user k as $E[T_k]$; the delay requirement of real-time user k can be formulated as

$$E[T_k] \leq \tau_k, \quad (1)$$

where τ_k is the delay bound of user k .

Denote the channel gain obtained by user k on subcarrier n by $h_{k,n}$ and the number of bits supported in a subcarrier by b . Define $v(k, n)$ to be an allocation indicator:

$$v(k, n) = \begin{cases} 1, & \text{if subcarrier } n \text{ is allocated to user } k, \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Our objective is to maximize the total system throughput, subject to the constraints on the total transmission power, user SNR requirements, and delay constraints. The optimization problem can be expressed as follows:

$$\max \sum_{k=1}^K \sum_{n=1}^N bv(k, n), \quad (3)$$

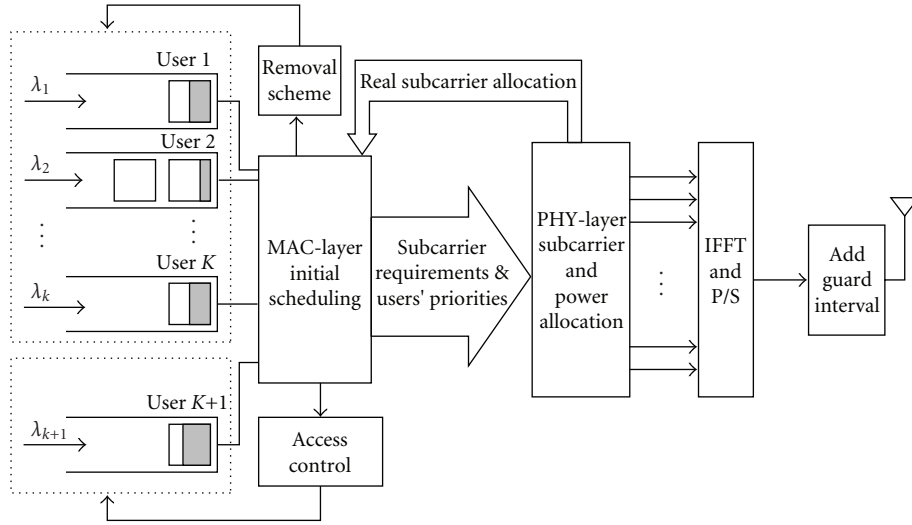


FIGURE 1: System model.

subject to

$$\begin{aligned}
 \text{C1: } & \sum_{k=1}^K \sum_{n=1}^N v(k, n) \leq N, \\
 \text{C2: } & \sum_{k=1}^K \sum_{n=1}^N \frac{\text{SNR}_k}{H_{k,n}^2} v(k, n) \leq P, \\
 \text{C3: } & v(k_1, n)v(k_2, n) = 0, \quad \forall k_1 \neq k_2 \in [1, K], \\
 \text{C4: } & E[T_k] \leq \tau_k, \quad \forall k \in [1, K],
 \end{aligned} \tag{4}$$

where SNR_k represents the SNR requirement of user k . C1 states that the total subcarriers allocated to all users are less than or equal to N ; C2 shows that the total transmission power should be less than or equal to the system power limit, while satisfying all users' SNR requirements; C3 means that no more than one user transmits in the same subcarrier; C4 is the average delay requirement of each user.

The solution of the above optimization problem (3) is not explicit due to the fact that C4 is not directly related to $v(k, n)$. Thus in the following section, we will establish the relationship between them and give the suboptimal subcarrier allocation solution $v(k, n)$ for each symbol with lower computational complexity.

3. Cross-Layer Algorithm Description

Based on queuing theory, the MAC-layer scheduling scheme is developed to calculate the users' transmission priorities and their corresponding specific bandwidth requirements in terms of the number of subcarriers. With the channel state information, users' SNR requirements and the system power constraints, the PHY-layer resource allocation scheme can deduce the maximum attainable throughput for each supported user. In addition, the MAC-layer access control and removal scheme will be triggered to adjust the number of users being served and provide the QoS guarantee for the active users in the system.

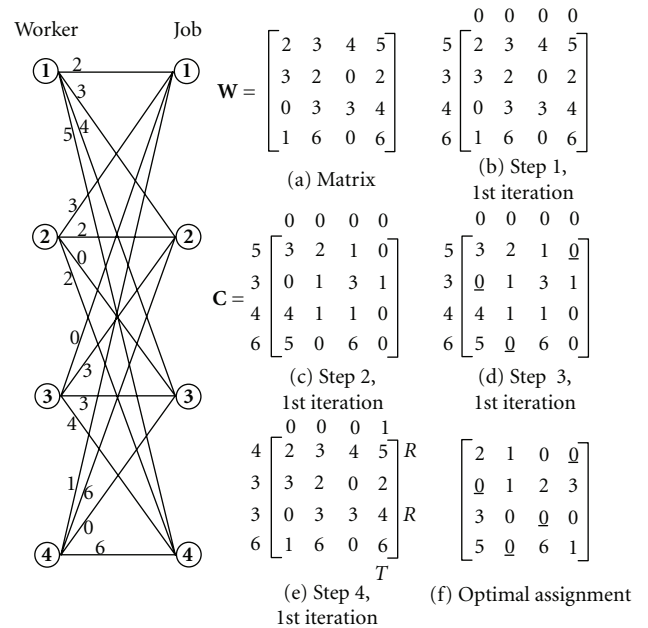


FIGURE 2: Weighted bipartite matching.

3.1. MAC-Layer Scheduling Scheme. In our system, we assume that each user has one type of real-time traffic. The packet arrivals of user k follow an independent Poisson process with rate λ_k , and each user has a delay upper bound τ_k . Furthermore, we assume that users have infinite buffers, and the same class users have the same (λ_k, τ_k) settings. Since the transmission process for each user can be modelled as an M/G/1 queue, the delay constraint on system time $E[T_k] \leq \tau_k$ is given by [11]

$$E[T_k] = E[X] + \frac{\lambda E[X^2]}{2(1-\rho)} \leq \tau_k, \tag{5}$$

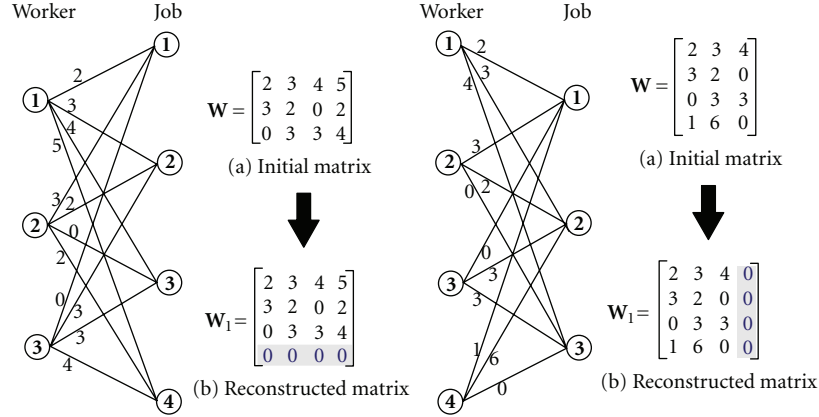
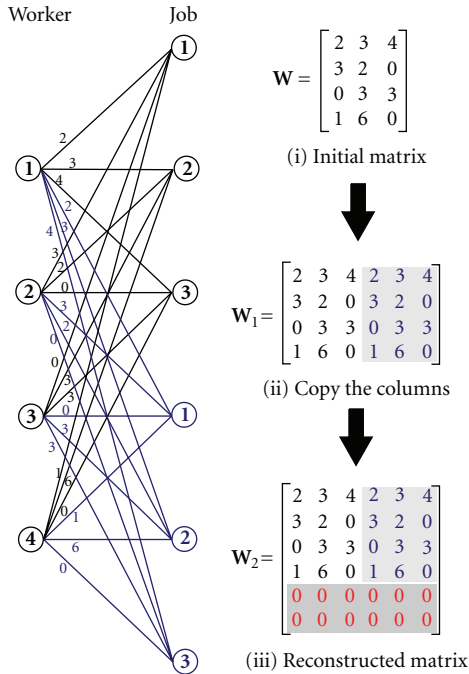


FIGURE 3: Asymmetric bipartite matching without resource reallocation.

FIGURE 4: Asymmetric bipartite matching with resource reallocation ($n_1 > n_2$).

where $E[X]$ is the average service time and $\rho = \lambda E[X]$. Since $E[X^2] = \text{Var}[X] + (E[X])^2 \geq (E[X])^2$, a necessary condition for the delay requirement on system time in (13) is

$$E[X] + \frac{\lambda(E[X])^2}{2(1-\rho)} \leq \tau_k. \quad (6)$$

By solving the above inequality, we can easily obtain the lower bound of the average transmission rate for user k . Since b is known by the supported modulation, we further scale the average transmission rate in terms of subcarriers, represented by R_k . Given the per-link R_k , the waiting time of the HOL packet w_k , and the delay constraint τ_k , an active user's transmission priority and exact bandwidth requirement in

terms of the number of subcarriers per symbol are obtained by the following modified LWDF scheduling algorithm.

In our algorithm, the system time is scaled in terms of OFDM symbol time. The remaining time to the deadline of the HOL packet at queue k is

$$r_k = \left\lceil \frac{\tau_k - w_k}{s} \right\rceil, \quad \forall k \in [1, K], \quad (7)$$

where s is the OFDM symbol time. The smaller the value of r_k is, the more urgently user k needs to transmit the corresponding packet. In addition, if l_k is the number of bits left in the HOL packet of user k , then till the due time of the packet, the average required transmission rate in terms of the number of subcarriers in the following symbol time is given by

$$Q_k = \left\lceil \frac{l_k}{br_k} \right\rceil, \quad \forall k \in [1, K]. \quad (8)$$

Compared with the deduced R_k , we define the rate proportional index as follows:

$$\zeta_k = \frac{R_k - Q_k}{R_k}, \quad \forall k \in [1, K]. \quad (9)$$

ζ_k is defined to indicate the urgent state. Its value could be positive or negative. If its value is below zero, this means that the required number of subcarriers exceeds the average number, which indicates that congestion may happen. It is also easily observed that the smaller the value of ζ_k , the more urgent the transmission of the corresponding HOL packet. As in LWDF algorithm, we also consider the factors of the waiting time and transmission rate for each user. However, instead of considering the users' attainable bandwidths, we consider the users' required bandwidths under the delay bound constraints, which are more important for real-time services. Our scheduling is described as follows. Once the channel is idle, each user will calculate its transmission priority by

$$\zeta_k r_k^{1/\alpha} \delta(N - Q_k), \quad \forall k \in [1, K], \quad (10)$$

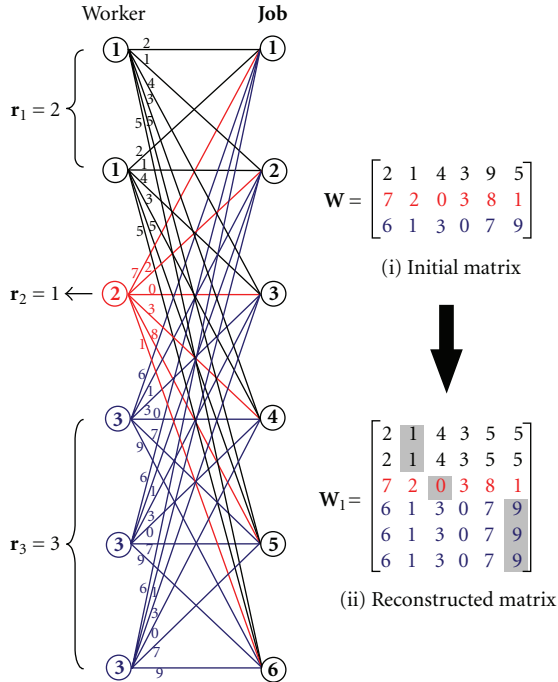


FIGURE 5: Asymmetric bipartite matching with resource reallocation ($n_1 < n_2$).

where α is a positive constant used to adjust the weight of r_k . The function $\delta(\cdot)$ is defined as

$$\delta(x) = \begin{cases} 1, & x \geq 0, \\ \infty, & x < 0. \end{cases} \quad (11)$$

From the above analyses, the user with the smaller value given by (10) will enjoy a higher transmission priority. From the definition of $\delta(\cdot)$, if a user's required number of subcarriers exceeds the total number N provided by a symbol, even if we allocate the whole symbol to this user, its delay requirement will not be met. Therefore, the HOL packet of this user will be dropped to save bandwidth for other users.

Up to now, our MAC-layer scheduling scheme gives the transmission priority list of the HOL packets according to (10) for the active users and their expected transmission rates in terms of the number of subcarriers in each symbol from (8). However these rates are only the users' expected rates. Considering the users' channel states, SNR requirements, and system power limit in the PHY layer, the real subcarrier allocation will be performed according to the following scheme.

3.2. PHY-Layer Resource Allocation Scheme. In the MAC layer, our algorithm has already considered the real-time traffic delay requirement and given the expected transmission rate in terms of the number of subcarriers and users' transmission priorities. In the PHY layer, with the different

subcarriers' channel states, system SNR and power constraints, our PHY-layer scheme aims to optimize the initial allocation indicator $v(k, n)$ with the following constraint:

$$\min \sum_{k=1}^K \sum_{n=1}^N \frac{\text{SNR}_k}{h_{k,n}^2} v(k, n) \leq P. \quad (12)$$

To solve this problem, a dynamic PHY-layer resource allocation scheme is proposed which is divided into the following steps.

- (a) *Initial subcarrier allocation.* With the total number of subcarrier limit N , we initially assign the users the required numbers of subcarriers according to their priorities till N subcarriers are used up or all K users are assigned.
- (b) *Power minimization.* Given a subcarrier allocation, the following modified Kuhn-Munkres algorithm is used to obtain an optimal allocation to minimize the system power under the users' SNR requirements. Denote the minimized power as P_{\min} .
- (c) *Power comparison.* Compare P_{\min} with the system power limit P , and consider the following cases:

- (i) if $P = P_{\min}$, then the power resource is fully utilized, and the current subcarrier allocation $v(k, n)$ is the final solution;
- (ii) if $P < P_{\min}$, then the system power cannot support all currently assigned subcarriers. So our scheme will reduce the subcarrier allocation from the lowest priority user. Given SNR_k requirement for user k , among the assigned subcarriers for this user, the smaller the value of $h_{k,n}$ on subcarrier n , the larger the power consumption on it. So the subcarrier reduction will be performed in ascending subcarrier gain order one by one. Then go to Step (b) in the next iteration, till the updated P_{\min} is less than P ;
- (iii) if $P > P_{\min}$, more power resource can be utilized. Then our scheme considers the remaining subcarrier resource. We represent the total number of the assigned subcarriers as N' . If $N' = N$, the subcarriers are used up, and we maintain the current $v(k, n)$ solution. If $N' < N$, the remaining subcarriers are assigned evenly to the current active users till the updated P_{\min} reaches P . If new users' access requirements are received, the access control scheme to be introduced in the next subsection will guide the assignment.

Modified Kuhn-Munkres Algorithm. In the following, we will firstly introduced the Kuhn-Munkres algorithm to find the perfect matching with the maximum sum of edge weights for a bipartite graph. Then a modified algorithm is described for OFDM power allocation. To minimize the system power, the modified algorithm is applied with negative weights.

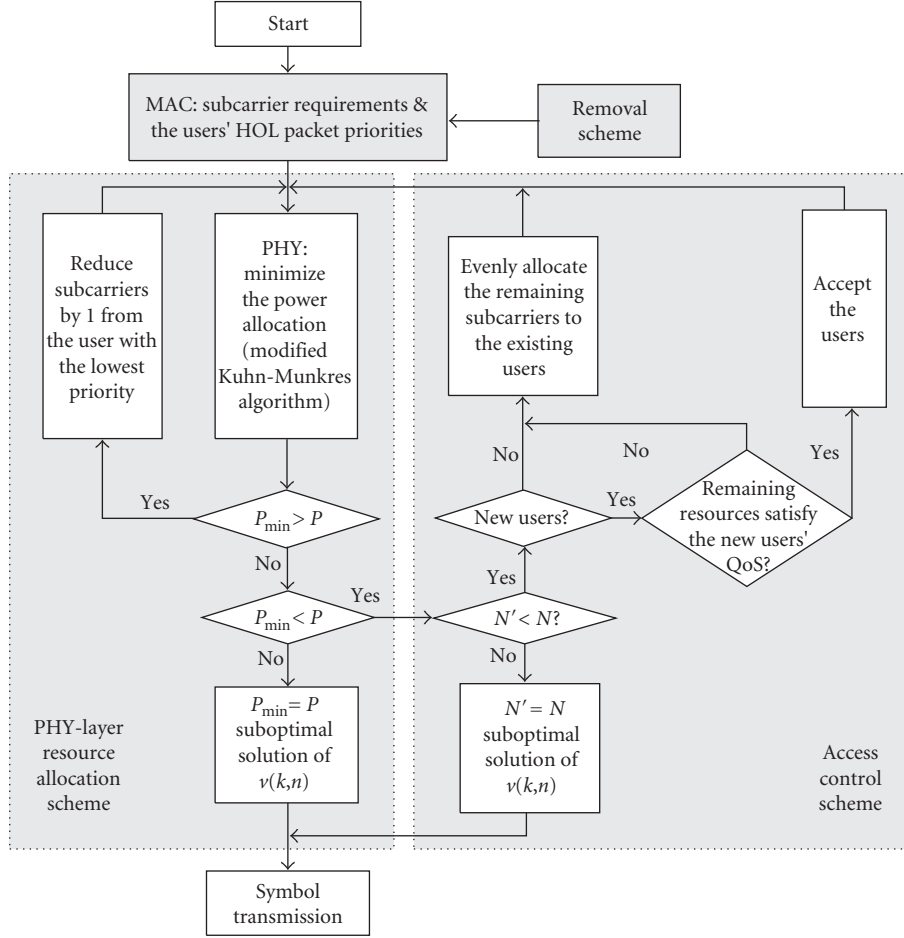


FIGURE 6: Flow chart of DSA-RT.

A graph is denoted by $G(V, E)$, where V is the vertex set, and E is the edge set of the graph. If $V = V_1 \cup V_2$ with $V_1 \cap V_2 = \Phi$ and each edge in E has one endpoint in V_1 and the other in V_2 , the graph $G(V, E)$ is a bipartite graph, which can also be denoted as $G(V_1, V_2, E)$. The bipartite graph is very useful for some applications, such as an assignment problem which can be depicted as follows. Given a weighted complete bipartite graph $G = (X \cup Y, X \times Y)$, where edge (x, y) has weight $w(x, y)$, find a matching m from X to Y with maximum weight. In an application, X could be a set of workers, Y a set of jobs, and $w(x, y)$ the earnings made by assigning worker x to job y . The goal of the assignment problem is to find the optimal (best total earnings) matching.

For a bipartite graph $G(V_1, V_2, E)$, if the cardinalities of V_1 and V_2 , denoted as n_1 and n_2 , are equal, then this bipartite graph is symmetric. For single objective optimization, it has been proved that the Kuhn-Munkres algorithm can always find the maximum weight matching for a bipartite graph with $O(n^3)$ computational complexity. The Kuhn-Munkres algorithm is based on the procedure of the Hungarian algorithm [9]. Matrix $W = [w_{ij}]$ has elements w_{ij} , which represent the earnings of assigning worker i to job j as shown in Figure 2 (a).

Step 1. Let X, Y be the bipartite sets. Initialize two labels u_i and v_j by $u_i = \max_j \{w_{ij}\}$, $v_j = 0, i, j = 1, \dots, k$. In Figure 2 (b), the numbers written at the left and the top of the matrix express the values of u_i and v_j , respectively.

Step 2. Obtain the excess matrix C by the following: $c_{ij} = u_i + v_j - w_{ij}$. This is shown in Figure 2 (c).

Step 3. Find the subgraph G' that includes vertices i and j satisfying $c_{ij} = 0$ and the corresponding edge e_{ij} . Then find the maximum matching m of G' by the Hungarian algorithm, and underline the entries in the weight matrix. (There are various ways to find the maximum matching. See, e.g., [12].) A maximum matching is a matching with the largest possible number of edges. In this example, the maximum matching is found to be $(1, 4)$, $(2, 1)$, and $(4, 2)$, as shown in Figure 2 (d). If m is a perfect matching, that is, the number of edges in a maximum matching is equal to the cardinality of worker set (k), the optimal assignment is obtained. Otherwise, go to the next step.

Step 4. Let Q be a vertex cover of G' , and let $R = X \cap Q$ and $T = Y \cap Q$. The vertex cover Q is a vertex set of G' which contains at least one endpoint of each edge. In this example,

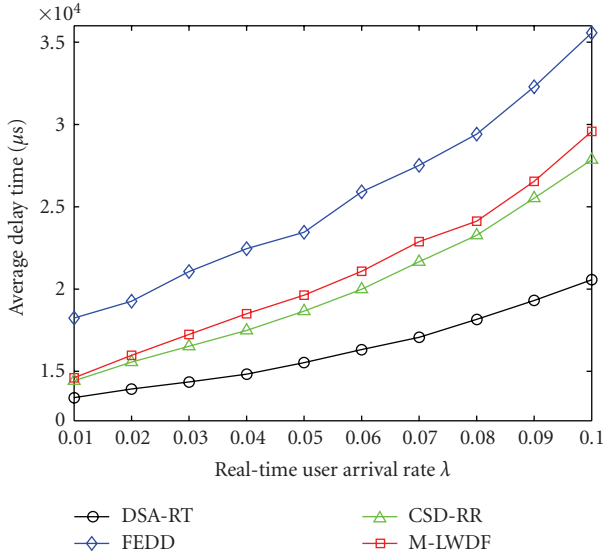


FIGURE 7: Average delay comparisons.

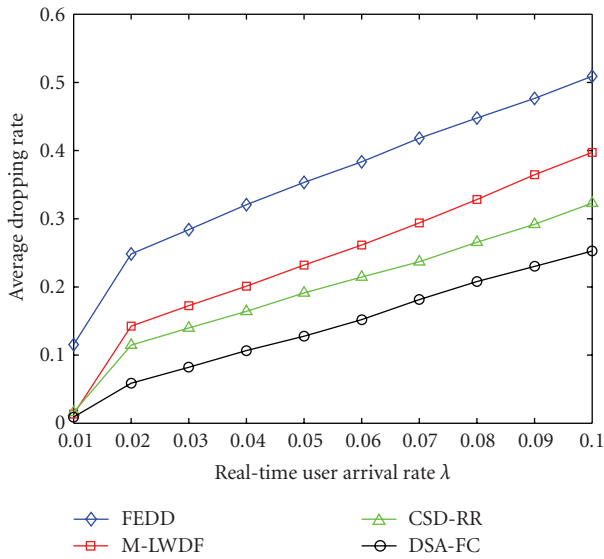


FIGURE 8: Dropping rate comparisons.

Q is chosen to be the nodes corresponding to Workers 1 and 3 and Job 4. So R corresponds to Workers 1 and 3, and T corresponds to Job 4. Now find $\epsilon = \min\{c_{ij} : x_i \in X - R, y_j \in Y - T\}$. For example, if ϵ equals 1 in Figure 2, decrease u_i by ϵ for the rows of $X - R$ and increase v_j by ϵ for the columns of T . Then go to Step 2.

Steps 2 to 4 are repeated until the perfect matching m , that is, the optimal assignment, is obtained.

For a bipartite graph $G(V_1, V_2, E)$, if the cardinalities of V_1 and V_2 , denoted as n_1 and n_2 , are not equal, then this bipartite graph is asymmetric. In our modified Kuhn-Munkres algorithm, we enhance an asymmetric graph to a symmetric one, and then solve the optimization problem as in the symmetric case. Firstly, suppose that the resource

on both V_1 and V_2 cannot be reused, we append $|n_1 - n_2|$ all-zero rows or columns to the weight matrix to construct a square matrix, and then transform the problem to a symmetric bipartite matching, as shown in Figure 3.

Secondly, for some special cases in which the redundant resource may be reused, the modified Kuhn-Munkres algorithm reproduces the corresponding columns or rows till the matrix is transformed to a square matrix. If necessary, all-zero columns or rows will be added. If $n_1 > n_2$ and the elements in V_2 is reusable, Figure 4 shows the case where the remaining elements in V_1 may reuse the elements in V_2 with the same probability. If $n_1 < n_2$, given the number of required elements in V_2 by the elements in V_1 , namely, q_1, q_2, \dots, q_{n_1} , then the square matrix may be constructed by reproducing the rows in demand, as shown in Figure 5.

In the downlink OFDM system model, as in previous work, channel state information (CSI) is assumed to be available at base stations (BSs). In a multiuser system with frequency-selective fading, each user may experience a different channel frequency response, which is related to its location. The total frequency bandwidth is divided into N orthogonal subchannels, and suppose there are currently K active users in the system. Assume that S_k is the subchannel set for user k , q_k is the cardinality of set S_k . The value of q_k is initially obtained from the MAC layer scheduling scheme and dynamically changed by the PHY allocation scheme. Therefore, for user k , the required transmission power in time slot t is given by

$$p_k(t) = \sum_{i=1}^{i=q_k} \frac{\text{SNR}_k}{h_{k,n}^2}, \quad (13)$$

where $h_{k,n}$ is the detected subchannel gain of user k on subchannel i . Then the total system required power can be expressed as

$$P(t) = \sum_{k=1}^K p_k(t) = \sum_{k=1}^K \sum_{i=1}^{i=q_k} \frac{\text{SNR}_k}{h_{k,n}^2}. \quad (14)$$

With the above problem formulation, the minimization of the system power $P(t)$ as required in the second step of the PHY-layer allocation scheme may be converted to a bipartite matching problem. The edge weight for user k on subcarrier n is $\text{SNR}_k/h_{k,n}^2$. Therefore, similar to the case illustrated in Figure 5, the modified Kuhn-Munkres algorithm may be applied to give an optimal solution to the minimization of the system power.

3.3. Access Control and Removal Scheme. In real networks, the number of active users changes dynamically. Without access control, the bandwidth may be inadequate. In addition, particularly for real-time traffic, without a removal scheme, not only may the QoS of the users newly granted access not be guaranteed but also the previously granted access users will suffer from QoS degradation. Therefore, the MAC-layer access control and removal schemes are introduced in our DSA-RT algorithm.

As analyzed in the previous subsection, a new user's QoS requirements should be considered when $P > P_{\min}$ and

$$N' = \sum_{k=1}^K \sum_{n=1}^N v(k, n) \leq N. \quad (15)$$

As introduced in Section 3.1, the new user's QoS requirements can be evaluated by R_k . Access control will check if this requirement can be satisfied with the remaining power and subcarrier resources. If yes, the new user can be allocated subcarrier resources; otherwise, it continues to wait.

Even with access control, real-time transmission systems may still encounter an overloaded situation due to the time-varying wireless channel and variable bit rates. As presented in [13], a useful removal scheme can effectively guarantee the QoS of the existing users and will not be adversely affected by the admission of new users. Our scheme assumes that the dropping rate of user k is sampled for each constant time interval Δt , and the last sample time is t . So the dropping rate of user k is defined by

$$\eta_k(t + \Delta t) = \frac{D_k(t + \Delta t)}{N_k(t + \Delta t)}, \quad (16)$$

where $D_k(t + \Delta t)$ and $N_k(t + \Delta t)$ are the numbers of dropped packets and the total transmitted packets of user k during time $(t, t + \Delta t]$. Assume θ_k is the maximum dropping rate which user k can tolerate. At each sample time or when the number of users in the system changes, our removal scheme will select the user to be removed by the following rule:

$$\{i\} = \underset{k}{\operatorname{argmax}} \left\{ \frac{\eta_k}{\theta_k} \right\}, \quad (17)$$

where the selected set consists of the users whose η_k values violate their corresponding dropping rate bound θ_k . If the traffic is bursty, we may change Δt to adjust the dropping rate more frequently.

3.4. Implementation of DSA-RT. Thanks to the cooperation of the above schemes, for each OFDM symbol, our algorithm DSA-RT can give the suboptimal solution $v(k, n)$ of the optimization problem addressed in Section 2. The computational delay is not expected to be a problem. The number of operations required by the algorithm is approximately $O(N^3)$, which translates to a computational delay of a small fraction of a symbol time with the support of current chips. In addition, if we want to lower the computational delay, multiple symbols can be combined as one scheduling unit, but this will affect the scheduling efficiency. It is a tradeoff. The flow chart of the implementation of our algorithm is shown in Figure 6.

4. Simulation Results

In this section, the performance of the proposed DSA-RT scheduling algorithm is investigated and compared with CSD-RR, FEDD, and M-LWDF [2–4]. We consider QPSK modulation in multiuser OFDM downlink systems.

However, other modulations are supported with different SNR constraints. The IFFT size is 128, and the OFDM symbol duration is equal to 200 microseconds [14]. We consider the quasistatic flat fading channel with multipath [15]. Assume that the users arrive as a Poisson process with parameter λ , and their active times in the system follow the exponential distribution with mean 10 seconds. In this section, we assume that all users have the same type of real-time traffic. During each user's active time, the packet arrivals follow the Poisson distribution. The packets have a fixed length of 1000 bytes, and the mean traffic rate is 1 Mbps. The delay bound is set to be 50 milliseconds. In simulations, we consider one type of real-time traffic, so we fixed the packet length. However, if multiple types of real-time traffics are supported, a variable length is acceptable in our algorithm. In our simulations, we vary the user arrival rates λ from 0.01 to 0.1 and compare the delay and dropping rate performance of some packet scheduling algorithms and our proposed DSA-RT algorithm. All simulations are in Matlab 7.3. The simulation time of each experiment is 100 seconds and we repeat it 100 times.

The average delay is the mean of the delay of all packets not dropped. For each successfully delivered packet, the delay is calculated as the difference between the departure and arrival times. In DSA-RT, packets which have been dropped will not re-enter the system. Figure 7 shows the delay comparisons of DAS-RT and three other packet scheduling algorithms. It is obvious that our algorithm distinctly improves the delay performance, particularly when the traffic density is high. Accordingly, as shown in Figure 8, the dropping rates of our algorithm at any user arrival rate are also much lower than the other three algorithms. DSA-RT is developed to schedule at the subcarrier level and tries to provide delay guarantees for the real-time traffics. Therefore, it has the best delay and dropping rate performance. Based on the consideration of channel state, CSD-RR has better performance than M-LWDF and FEDD. By considering the system capacity and queuing, the throughput performance of M-LWDF is optimal, but the delay performance still needs to be improved. FEDD gives the packet with the earliest deadline of the highest transmission priority. However, with the bandwidth and channel state constraints, the transmission still has a high probability to fail within its deadline. Therefore, it has the poorest performance.

5. Conclusion

In this paper, DSA-RT aims to satisfy the packet delay requirements of real-time traffics in multiuser OFDM system, while maximizing the system bandwidth efficiency. This algorithm consists of two cooperative components. At the MAC layer, based on queuing theory and the modified LWDF algorithm, active users' expected transmission rates in terms of the number of subcarriers per symbol and their corresponding transmission priorities are deduced. With different subcarrier states, based on our modified Kuhn-Munkres algorithm, a PHY-layer resource allocation scheme is developed to satisfy the users' requirements under the

system SNR and power constraints. When considering a system where the number of active users changes dynamically, the access control and removal scheme can fully utilize the bandwidth resource and guarantee the QoS of the existing users in the system. Finally, compared with other widely used scheduling algorithms, simulation results show that our proposed algorithm significantly improves the system performance for real-time users in multiuser OFDM systems.

References

- [1] E. Lawrey, "Multiuser OFDM," in *Proceedings of the International Symposium on Signal Processing and Its Applications (ISSPA '99)*, pp. 761–764, Brisbane, Australia, August 1999.
- [2] P. Bhagwat, A. Krishna, and S. Tripathi, "Enhancing throughput over wireless LAN's using channel state dependent packet scheduling," in *Proceedings of 17th Annual Joint Conference of the IEEE Computer and Communications Society (INFOCOM '98)*, pp. 1103–1111, San Francisco, Calif, USA, March 1998.
- [3] S. Shakkottai and R. Srikant, "Scheduling real-time traffic with deadlines over a wireless channel," *Wireless Networks*, vol. 8, no. 1, pp. 13–26, 2002.
- [4] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Communications Magazine*, vol. 39, no. 2, pp. 150–153, 2001.
- [5] Y. J. Zhang and S. C. Liew, "Link-adaptive largest-weighted-throughput packet scheduling for real-time traffics in wireless OFDM networks," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '05)*, vol. 5, pp. 2490–2494, St. Louis, Mo, USA, November–December 2005.
- [6] Y. Y. W. Cheong, C. Y. Tsui, R. S. Cheng, and K. B. Letaief, "A realtime sub-carrier allocation scheme for multiple access downlink OFDM transmission," in *Proceedings of the 49th IEEE Vehicular Technology Conference (VTC '99)*, vol. 2, pp. 1124–1128, Amsterdam, The Netherlands, September 1999.
- [7] Z. Diao, D. Shen, and V. O. K. Li, "CPLD-PGPS scheduler in wireless OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 5, no. 10, pp. 2923–2931, 2006.
- [8] J. Munkres, "Algorithms for the assignment and transportation problems," *Journal of the Society for Industrial and Applied Mathematics*, vol. 5, no. 1, pp. 32–38, 1957.
- [9] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistic Quarterly*, vol. 2, pp. 83–97, 1955.
- [10] J. Zhu, B. Bing, Y. Li, and J. Xu, "An adaptive subchannel allocation algorithm for OFDM-based wireless home networks," in *Proceedings of the 1st IEEE Consumer Communications and Networking Conference (CCNC '04)*, pp. 352–356, Las Vegas, Nev, USA, January 2004.
- [11] D. Bertsekas and R. Gallager, *Data Networks*, Prentice-Hall, Englewood Cliffs, NJ, USA, 2nd edition, 1992.
- [12] D. B. West, *Introduction to Graph Theory*, Prentice-Hall, Englewood Cliffs, NJ, USA, 2001.
- [13] E. Kwon, S.-G. Kim, and J. Lee, "Overload control with removal algorithm for real-time flows in wireless networks," in *Proceedings of the IEEE 63rd Vehicular Technology Conference (VTC '06)*, vol. 3, pp. 1127–1131, Melbourne, Australia, May 2006.
- [14] J. Cai, X. Shen, and J. W. Mark, "Downlink resource management for packet transmission in OFDM wireless communication systems," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '03)*, vol. 6, pp. 2999–3003, San Francisco, Calif, USA, December 2003.
- [15] C. Xiao, Y. R. Zheng, and N. C. Beaulieu, "Second-order statistical properties of the WSS Jakes' fading channel simulator," *IEEE Transactions on Communications*, vol. 50, no. 6, pp. 888–891, 2002.