Cross-Layer Joint Rate Control and Scheduling for OFDMA Wireless Mesh Networks

Mohammad Fathi, *Student Member, IEEE,* Hassan Taheri, *Member, IEEE,* and Mehri Mehrjoo, *Member, IEEE*

***Abstract*—We consider the utility maximization problem in the downlink of wireless mesh networks (WMN) with orthogonal frequency division multiple access (OFDMA). We formulate this problem as a cross-layer design of joint rate control and OFDMA scheduling in order to utilize the scarce radio resources efficiently. The problem is decoupled into a rate control problem, at the transport layer, and a channel-aware and queue-aware scheduling problem, at the MAC/PHY layer. The rate control problem adjusts arrival rates to the base station (BS) queues, and the scheduling problem, determines link rates, i.e., departure rates from the BS and other network nodes, through subcarrier and modulation rate assignment. While the rate control problem is solved locally at the BS, we propose a greedy algorithm that solves the scheduling problem in a distributed manner, at network nodes. Furthermore, we propose a heuristic algorithm for fast execution of the scheduling scheme at individual nodes. Numerical results show that the heuristic algorithm presents a comparable performance to that of the greedy algorithm, while it has lower computational complexity. Besides, our proposed scheduling scheme, cooperating with the rate control mechanism, improves the network performance in terms of end-to-end delay, aggregate utility, and fairness.**

***Index Terms*—Wireless mesh networks, cross-layer design, rate control, resource allocation, scheduling, OFDMA, decomposition, and optimization.**

I. INTRODUCTION

Wireless mesh networking is an emerging technology for scalable broadband Internet access [1]. Wireless mesh networks (WMN) have been considered to be implemented in several standards including IEEE 802.11 and IEEE 802.16 [2]. A common type of WMN consists of a base station (BS) and several intermediate nodes called relay nodes (RNs) which provide routes between the BS and end-nodes. In general, RNs are fixed and connected to the main power supply, so they can improve network capacity and scalability. To fully realize WMN advantages, research works are carried out toward the network performance optimization [3].

A common approach in WMNs performance optimization is network utility maximization (NUM) through cross-layer design of control mechanisms and resource allocation schemes. Cross-layer design takes users' service requirements, shared and interference-limited channel into account to jointly optimize the network performance. Better network performance can be obtained by cross-layer resource allocation, which may not be achieved by traditional layering architecture [4]. In fact, the link capacities determined by the resource allocation at the MAC/PHY layer influence arrival rates at the transport layer and vice versa. In other words, a rate control mechanism cannot be effective, unless resource allocations at underlying layers support it. With this coupling, resource optimization within layers is not competent, and a cross-layer approach should be employed to achieve the optimal performance. Therefore, a cross-layer NUM problem is proposed to comprise performance optimization of the transport layer and MAC/PHY layer jointly [5]–[15], as shown in Fig. 1.

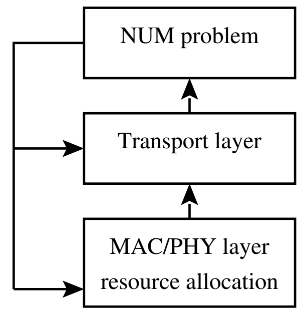


Fig. 1. A cross-layer NUM problem diagram

Kelly, in his seminal work [16] presented a framework for NUM in communication networks. Motivated by this framework, cross-layer design of NUM has been extensively studied in single carrier wireless networks [5]–[15]. In these works, dual decomposition is used to decouple the problem into different functional modules in the network protocol stack. A joint rate control and centralized scheduling is proposed for contention-based wireless ad hoc networks with constant channel in [5]–[7]. A distributed algorithm for the scheduling problem in these works is presented in [8] by considering only primary interference. In [9], [10], joint centralized scheduling and power control are presented for ad hoc networks subject to a minimum SINR threshold on each link. A joint congestion control and centralized scheduling for multi-hop wireless networks with constant channel states has been presented in [11]. Due to the high complexity of the scheduling solution, the authors have investigated the impact of an imperfect but fast implementable scheduling on the cross-layer congestion control in [12]. It has been shown that this approach substantially outperforms layered approaches which do not consider cross-layer design. In [13], the stability of a stochastic control policy,

M. Fathi and H. Taheri are with the Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran (email: {mfathi, htaheri}@aut.ac.ir).

M. Mehrjoo is with the Faculty of Electrical and Computer Engineering, University of Sistan and Baluchestan, Iran (email: mehrjoo@ece.usb.ac.ir).

This work was supported in part by Iran Telecommunication Research Center (ITRC).

which is decoupled into separate algorithms for rate control, routing, and centralized scheduling, is provided for networks with time-varying channels. The asymptotic stability of a primal-dual congestion controller supported by a centralized scheduling is proved for constant and time varying channels in [14] and [15], respectively. Cross-layer design in ad hoc networks using convex optimization has been surveyed in [17].

In the above literature, capacity region is assumed to be convex by allowing time sharing between independent sets of the conflict graphs [18] of the single carrier networks. Then, a centralized or distributed scheduling is proposed, respectively, assuming constant channel states or time-varying channels with simple interference models.

The key contribution of this paper is to propose a partially distributed scheduling policy for the NUM problem in multicarrier WMNs with time-varying channels. The capacity region is discrete and thus non-convex as a result of considering a finite set of modulation rates on each carrier. The multicarrier technique is Orthogonal Frequency Division Multiplexing (OFDM) which divides a broadband channel into a set of non-interfering narrowband subcarriers with independent channel gains. In a multiuser OFDM network, different subcarriers can be allocated to different users to provide a multiple access method denoted as Orthogonal Frequency Division Multiple Access (OFDMA). Employing OFDMA improves WMN performance by adaptively allocating subcarriers and transmission power based on instantaneous channel state information (CSI).

While single-hop OFDMA resource allocation has been extensively studied, a few works exist for the multi-hop case. In the single-hop scheduling algorithm proposed in [19]–[23], assuming fixed power allocation, subcarriers are assigned dynamically, and then a greedy power allocation algorithm is employed for bit loading. For multi-hop OFDMA, network capacity maximization subject to a minimum rate provisioning at intermediate nodes has been investigated in [24]–[27]. In [24], a two-level distributed hierarchical scheduling (DHS) has been proposed. First, a mesh router determines the number of subcarriers to be assigned to each intermediate node called mesh client. Second, each mesh client performs subcarrier assignment and power allocation to its outgoing links. In both steps, the problem is formulated by nonlinear integer programming that is solved by integer relaxation. In [25], an opportunistic subchannel and power scheduling algorithm has been proposed for both BS and RNs. Considering a two-hop relay network in [26], the resource allocation problem has been solved by separating the subcarrier assignment and power allocation.

In this paper, we formulate the joint rate control and OFDMA resource allocation of multicarrier WMNs by a NUM problem. Using dual decomposition, the problem is decoupled into a rate control problem and a multi-hop OFDMA scheduling problem at the transport layer and MAC/PHY layer, respectively. The solution of the rate control problem adjusts queues arrival rates at the BS. On the other hand, the scheduling problem determines link rates by subcarrier and modulation rate assignment throughout the network. We formulate the scheduling problem as an integer programming problem and propose a greedy algorithm that solves it in a partially distributed manner. For fast execution of the scheduling scheme, we further propose a heuristic algorithm that speeds up the operations.

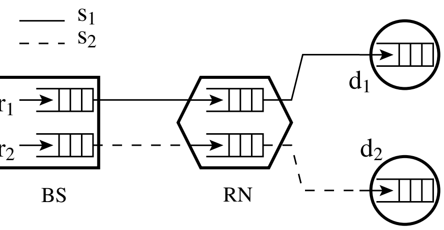


Fig. 2. A two-hop network architecture

The rest of the paper is organized as follows. Network model including the network architecture and radio transmission model along with the problem formulation are described in Section II. The NUM problem is decomposed into the rate control and scheduling problems in Section III. In Sections IV and V, the greedy and heuristic solutions for the scheduling problem are presented, respectively. Performance evaluation results are given in Section VI, and the paper is concluded in Section VII.

II. Network model and probelm formulation

In this section, network architecture and radio transmission model are described in subsection A, and NUM problem is formulated in subsection B. The following assumptions are made throughout the paper: a) We consider a time slotted transmission where CSI remains valid within a time slot but varies randomly and independently across time slots. b) Since topology doesn't change frequently in WMNs, each flow is routed along a fixed path from the BS to the corresponding destination. c) Each RN is equipped with two radio interfaces for simultaneous transmission and reception. Therefore, it can transmit on one subcarrier and, at the same time, can receive on another subcarrier. d) Subcarriers are not shared, i.e., each subcarrier is exclusively assigned to one link at a time.

A. Network Architecture and Radio Transmission Model

We consider a WMN with a set $\Phi = \{s : s = 1, 2, ..., S\}$ of flows transmitted from the source node, the BS, to corresponding destination nodes, $d_s : s \in \Phi$. Each flow s is transmitted with the rate r_s along the path L_s consisting of a set of links (denoted as $ij \in L_s$) on the route from the BS to d_s. The arrival flows to the intermediate nodes are buffered in separate queues. As an example, Fig. 2 illustrates a two-hop network in which the BS sets two flows s_1 and s_2 via a RN to the corresponding destinations. Let $C = \{c_{ij}\}$ denotes a link capacity vector, where c_{ij} is the capacity of link ij (the link from transmitting node i to receiving node j). The link capacity vector is determined by the scheduling scheme which assigns a set $\Omega = \{k : k = 1, 2, ..., K\}$ of OFDM subcarriers to links and allocates node i's transmission power, P_i, to subcarriers. The scheduling decision is made periodically at the beginning of each downlink interval containing a number of OFDM symbols, denoted as time slots.

Signal to noise ratio (SNR) of subcarrier k on link ij, during a time slot, is equal to $h_{ij}^k p_{ij}^k$, where p_{ij}^k is the portion of P_i allocated to subcarrier k. Moreover, $h_{ij}^k = \frac{|H_{ij}^k|}{N}$, where N represents noise power density, and H_{ij}^k is the channel gain which is depending on path loss, shadowing, and fading. We consider adaptive rate allocation with a finite set of modulation rates $D = \{0, 1, ..., M\}$, where M is the highest modulation rate. Accordingly, the number of transmitted bits on link ij and subcarrier k, c_{ij}^k, is given by

$$c_{ij}^k = \min\{\left\lfloor \log_2\left(1 + h_{ij}^k p_{ij}^k\right)\right\rfloor, M\} \text{ bps/Hz.} \tag{1}$$

B. NUM Problem

We define the NUM problem parameters as follows: Let $R = \{r_{BS}^{(s)} \geqslant 0 : s \in \Phi\}$ be the set of long-term average arrival rates to the BS, $\Psi = \{C : C = [c_1, c_2, ..., c_{ij}, ...]\}$ be the set of all feasible link capacity vectors. Also $F = \{f_{ij}^{(s)} \geqslant 0 : s \in \Phi, ij \in L_s\}$ is the set of link rates, where $f_{ij}^{(s)}$ denotes the capacity portion of link ij that is allocated to flow s.

Assume that each flow s is associated with a utility function U_s which is continuously differentiable, non-decreasing, and strictly concave for elastic traffic. The objective of the NUM problem is to maximize the sum of utilities, functions of arrival rates to the BS, subject to the network constraints at different layers. We formulate the NUM problem as in the following:

$$\text{P1}: \quad \max_{R,\,F,C} \sum_{s:n=BS} U_s(r_n^{(s)}) \tag{2}$$

$$\text{s.t.} \quad r_n^{(s)} + \sum_{i:in\in L_s} f_{in}^{(s)} \leqslant \sum_{j:nj\in L_s} f_{nj}^{(s)}, \ \forall s, n \neq d_s \tag{3}$$

$$f_{ij} = \sum_{s:ij\in L_s} f_{ij}^{(s)} \leqslant c_{ij}, \ \forall ij \tag{4}$$

$$f_{ij}^{(s)} = 0 \ \text{ if } ij \notin L_s, \ \forall s, ij \tag{5}$$

$$C \in \Psi. \tag{6}$$

Inequality (3) ensures that total rate s arriving in a non-destination node n is less than or equal to the total rate s out of this node. Note that $r_n^{(s)} = 0$ if $n \neq BS$, and $\sum_{i:in\in L_s} f_{in}^{(s)} = 0$ if $n = BS$. Constraint (4) states that the total rate of the link ij should not exceed the link capacity c_{ij}. Constraint (5) indicates that the rate s is zero on the link ij, if its path does not include this link. Finally, constraint (6) forces the link capacity vector to lie in the feasible region.

III. NUM PROBLEM DECOMPOSITION

We use dual decomposition to solve P1. Introducing $\Lambda = \{\lambda_n^{(s)} \geqslant 0, \text{ for all } n, s : n \neq d_s\}$, the set of Lagrange multipliers for constraint (3), Lagrangian function is given as

$$L(R, F, C, \Lambda) = \sum_{s:n=BS} U_s(r_n^{(s)}) - \sum_{s,n:n\neq d_s} \lambda_n^{(s)} \left(r_n^{(s)} + \sum_{i:in\in L_s} f_{in}^{(s)} - \sum_{j:nj\in L_s} f_{nj}^{(s)} \right). \tag{7}$$

Also, the dual function is defined as follows:

$$D(\Lambda) = \sup_{R,F,C} \{L(R, F, C, \Lambda) : R, F \geqslant 0, (4)\text{-}(6)\}. \tag{8}$$

Considering $r_n^{(s)} = 0$ for $n \neq BS$, we rewrite the dual function as

$$D(\Lambda) = \sup_{R\geqslant 0}\{ \sum_{s:n=BS} \left(U_s(r_n^{(s)}) - \lambda_n^{(s)} r_n^{(s)} \right)\} + \sup_{F,C\geqslant 0}\{ \sum_{s,n:n\neq d_s} \lambda_n^{(s)} \left(\sum_{j:nj\in L_s} f_{nj}^{(s)} - \sum_{i:in\in L_s} f_{in}^{(s)} \right) : (4)\text{-}(6)\}. \tag{9}$$

Due to the discrete capacity vectors, P1 is not convex. When this problem is solved in the dual domain, the duality gap decreases as the number of subcarriers increases [28]. In practice, the number of subcarriers is sufficiently large, so P1 can be solved in the dual domain efficiently. The corresponding dual problem is given as

$$\text{P2} : \min_{\Lambda\geqslant 0} D(\Lambda). \tag{10}$$

Using the dual function in (9) to evaluate $D(\Lambda)$ for a given Λ, we obtain the following optimization problems:

$$\text{P3} : \max_{R\geqslant 0} \sum_{s,n=BS} \left(U_s(r_n^{(s)}) - \lambda_n^{(s)} r_n^{(s)} \right), \tag{11}$$

and

$$\text{P4} : \max_{C,F\geqslant 0} \sum_{s,n:n\neq d_s} \lambda_n^{(s)} \left(\sum_{j:nj\in L_s} f_{nj}^{(s)} - \sum_{i:in\in L_s} f_{in}^{(s)} \right) \tag{12}$$

s.t. (4)-(6).

As seen, P3 optimizes arrival rates at the BS, and P4 optimizes link rates. Accordingly, they are called the rate control problem and the scheduling problem, respectively.

Given the solutions of P3 and P4, we use subgadient method to solve P2. Starting with an initial $\lambda_n^{(s)}(0)$ for all s and n, at each time slot t with a given $\lambda_n^{(s)}(t)$, the optimal value of rates $r_n^{(s)}(t)$ and link rates $f_{ij}^{(s)}(t)$ are obtained from P3 and P4, respectively. Then each $\lambda_n^{(s)}$ is updated by

$$\lambda_n^{(s)}(t+1) = \left[\lambda_n^{(s)}(t) - \kappa \left(\sum_{j:nj\in L_s} f_{nj}^{(s)}(t) - r_n^{(s)}(t) - \sum_{i:in\in L_s} f_{in}^{(s)}(t) \right) \right]^+, \tag{13}$$

where $(\sum_{j:nj\in L_s} f_{nj}^{(s)} - r_n^{(s)} - \sum_{i:in\in L_s} f_{in}^{(s)})$ is the subgradient of $D(\Lambda)$ with respect to $\lambda_n^{(s)}$. Moreover, the step size $\kappa > 0$ is chosen small enough to ensure the convergence [29].

Since $(r_n^{(s)} + \sum_{i:in\in L_s} f_{in}^{(s)})$ and $\sum_{j:nj\in L_s} f_{nj}^{(s)}$ denote inflow and outflow rates corresponding to flow s at node n, respectively, it is deduced that $\lambda_n^{(s)}$ is proportional to $Q_n^{(s)}$, the queue-length s at node n, i.e., $\lambda_n^{(s)} = \kappa Q_n^{(s)}$. We will use this derivation in Section IV. The above mentioned decomposition method is summarized in Algorithm 1.

Having Λ at each time slot t, we solve P3 and P4. Problem P3, which is solved only in the BS, can be decomposed into

Algorithm 1 Decomposition Algorithm

Input: CSI of each time slot.
Output: $r_n^{(s)}$, $\sum\limits_{i:in\in L_s} f_{in}^{(s)}$, and $\sum\limits_{j:nj\in L_s} f_{nj}^{(s)}$ for all n, s.
set $t=0$ and initialize $\lambda_n^{(s)}(t)$ for all n and s.
loop
  for all n and s **do**
    given $\lambda_n^{(s)}(t)$, obtain $r_n^{(s)}(t)$, $\sum\limits_{i:in\in L_s} f_{in}^{(s)}(t)$, and $\sum\limits_{j:nj\in L_s} f_{nj}^{(s)}(t)$ from $P3$ and $P4$.
    using (13), update $\lambda_n^{(s)}$ for the next time slot.
    return $r_n^{(s)}(t)$, $\sum\limits_{i:in\in L_s} f_{in}^{(s)}(t)$, and $\sum\limits_{j:nj\in L_s} f_{nj}^{(s)}(t)$.
  end for
  $t = t+1$.
end loop

subproblems of rate control for separate flows. Given $r_n^{(s)}(t)$, each subproblem s is solved using subgradient method as

$$\begin{aligned} &r_n^{(s)}(t+1)\\ &=\left[r_n^{(s)}(t)+\kappa\left(V\acute{U}_s(r_n^{(s)}(t))-\lambda_n^{(s)}(t)\right)\right]^+ \text{ for all } s, n=BS, \end{aligned} \tag{14}$$

where $(\acute{U}_s(r_n^{(s)})-\lambda_n^{(s)})$ is the subgradient of P3 objective function with respect to $r_n^{(s)}$, and $\kappa>0$ is the step size. Moreover, V is a constant that determines how aggressively this controller reacts to the same queue-length levels. The solution of scheduling problem P4 will be presented in Section IV.

IV. Scheduling Problem Solution

First, we reformulate the scheduling problem to specify MAC/PHY layer resource constraints in subsection A, then we present the solution in subsection B.

A. Scheduling Problem Formulation

Substituting $\lambda_n^{(s)}$ with $\kappa Q_n^{(s)}$, we rewrite the objective function in P4 as

$$\begin{aligned} &\max_{C,F\geqslant 0} \sum_{s,n:n\neq d_s} Q_n^{(s)}\left(\sum_{j:nj\in L_s} f_{nj}^{(s)}-\sum_{i:in\in L_s} f_{in}^{(s)}\right)\\ &=\max_{C,F\geqslant 0} \sum_{ij}\sum_{s:ij\in L_s} f_{ij}^{(s)}\left(Q_i^{(s)}-Q_j^{(s)}\right)\\ &=\max_{C,F\geqslant 0} \sum_{ij} f_{ij}^{(s^*)}\left(Q_i^{(s^*)}-Q_j^{(s^*)}\right). \end{aligned} \tag{15}$$

We have transformed the formulation from node centric to link centric in the first equality. In the second equality, we use the evidence that available rate on link ij, f_{ij}, is to be allocated to flow $s^*=\arg\max\limits_{s:ij\in L_s}(Q_i^{(s)}-Q_j^{(s)})$, i.e., $f_{ij}^{(s)}=f_{ij}$ if $s=s^*$, otherwise $f_{ij}^{(s)}=0$.

Under the assumption of backlog flows, link capacity provided by the scheduler on link ij, is fully utilized by flow s^*, i.e., $f_{ij}^{(s^*)}=c_{ij}$. Therefore, objective function in (15) can be replaced with

$$\max_C \sum_{ij} c_{ij}Q_{ij}, \tag{16}$$

where $Q_{ij}\equiv(Q_i^{(s^*)}-Q_j^{(s^*)})$ is the differential backlog on link ij. The objective function is a weighted sum capacity maximization problem. It is dynamic back pressure (DBP) [30] scheduling which allocates larger capacities to the links with larger differential backlogs.

In the following, we derive OFDMA capacity region constraints of the scheduling problem. We consider adaptive modulation with a finite set of modulation rates, $m\in D$. Therefore, link capacity c_{ij} is given as the sum of modulation rates corresponding to subcarriers assigned to link ij. The capacity is obtained by $c_{ij}=\sum\limits_k\sum\limits_m m\rho_{ij}^{(k,m)}$, where $\rho_{ij}^{(k,m)}$ is a binary variable equals to 1 if subcarrier k is exclusively assigned to link ij with modulation rate m , and 0 otherwise.

To avoid inter-link interference, each subcarrier is assigned exclusively to one link, i.e., $\sum\limits_{ij}\sum\limits_m \rho_{ij}^{(k,m)}=1$ for all k. Also, the total allocated power to the subcarriers assigned to a transmitting node i should not exceed the total power P_i, i.e., $\sum\limits_j\sum\limits_k\sum\limits_m p_{ij}^k\rho_{ij}^{(k,m)}\leqslant P_i$. Let subcarrier k be assigned to link ij with modulation rate m. From (1), we have $m\leqslant\log_2(1+h_{ij}^kp_{ij}^k)$ which implies $\frac{2^m-1}{h_{ij}^k}\leqslant p_{ij}^k$, so the power constraint at node i can be stated as $\sum\limits_j\sum\limits_k\sum\limits_m(\frac{2^m-1}{h_{ij}^k})\rho_{ij}^{(k,m)}\leqslant P_i$. Accordingly, the discrete rate multi-hop OFDMA scheduling problem is represented as follows:

$$\text{P5}:\quad \max_\rho \sum_{ij}\sum_k\sum_m m\rho_{ij}^{(k,m)}Q_{ij} \tag{17}$$

$$\text{s.t.}\quad \sum_{ij}\sum_m \rho_{ij}^{(k,m)}=1,\quad \forall k \tag{18}$$

$$\rho_{ij}^{(k,m)}\in\{0,1\},\quad \forall m,k,ij \tag{19}$$

$$\sum_j\sum_k\sum_m \rho_{ij}^{(k,m)}\left(\frac{2^m-1}{h_{ij}^k}\right)\leqslant P_i,\ \forall i. \tag{20}$$

Problem P5 is a binary integer programming problem with high complexity [31]. Complexity of this problem grows with the number of flows, subcarriers, and modulation rates. An optimal solution requires a centralized computation with exhaustive search that is clearly impractical in the large-scale networks. Hence we propose a partially distributed greedy algorithm, explained in subsection B. Using distributed scheduling, the processing load is distributed among the network nodes, which is highly desired in multi-hop communication networks.

B. Scheduling Problem Solution

The objective in P5 is to assign the subcarriers to the network links and to determine the modulation rate of each subcarrier. To reduce the difficulty, we propose a greedy algorithm that solves the problem partially distributed in two levels. First, in the BS, subcarriers are divided into distinct

Algorithm 2 Greedy algorithm: subcarrier assignment

1: **Input**: CSI of each time slot.
2: **Output**: Ω_i for all i.
3: initialize $\Omega_i = \emptyset$ for all i.
4: obtain $\bar{h}_i^k = E\{h_{ij}^k\}$ for all i and k.
5: **for all** k **do**
6: compute $m_i^k = \log_2(1 + \frac{\bar{h}_i^k P_i}{|\Omega_i|+1})$ for all i.
7: $i^* = \arg\max\limits_i (Q_i m_i^k)$.
8: $\Omega_{i^*} = \Omega_{i^*} \cup \{k\}$.
9: **end for**
10: return Ω_i for all i.

sets and each set is assigned to a transmitting node. Then, at each individual node, link and modulation rates are assigned to the corresponding subcarriers.

The first part of the greedy algorithm, subcarrier assignment to the network nodes, is proposed in Algorithm 2. In the Algorithm, Ω_i denotes the set of subcarriers assigned to node i, $|\Omega_i|$ is the cardinality of Ω_i, and $Q_i \equiv \sum\limits_j Q_{ij}$. At each iteration k in step 5, each node i allocates the same transmission power, $\frac{P_i}{|\Omega_1|+1}$, to the subcarriers which have already been assigned, in order to obtain allocated rate m_i^k. In step 7, $i^\star$ is the node for which assigning the subcarrier k results in the highest (rate $\times$ queue-length) in accordance with the objective function in P5. Having assigned subcarrier k to $i^\star$, subcarrier set $\Omega_{i^\star}$ is updated. Notice that m_i^k is an auxiliary parameter for subcarrier assignment, not the final and desired allocated rate.

In the sequel, we proceed with solving the scheduling problem in node i. Given Ω_i, we assigned the contained subcarriers to outgoing links ij's and determining corresponding modulation rates as formulated in problem P6:

$$\text{P6}: \quad \max_{\rho:k\in\Omega_i} \sum_j \sum_{k\in\Omega_i} \sum_m m\rho_{ij}^{(k,m)} Q_{ij} \tag{21}$$

$$\text{s.t.} \quad \sum_j \sum_m \rho_{ij}^{(k,m)} = 1, \ \forall k \in \Omega_i \tag{22}$$

$$\rho_{ij}^{(k,m)} \in \{0,1\}, \ \forall m, k, j \tag{23}$$

$$\sum_j \sum_{k\in\Omega_i} \sum_m \rho_{ij}^{(k,m)} \left(\frac{2^m - 1}{h_{ij}^k}\right) \leqslant P_i. \tag{24}$$

To solve P6, we form the Lagrangian function as

$$\begin{aligned} L_i(\rho, \mu_i) &= \sum_j \sum_{k\in\Omega_i} \sum_m m\rho_{ij}^{(k,m)} Q_{ij} \\ &- \mu_i \left(\sum_j \sum_{k\in\Omega_i} \sum_m \rho_{ij}^{(k,m)} \left(\frac{2^m - 1}{h_{ij}^k}\right) - P_i \right) \qquad (25) \\ &= \sum_j \sum_{k\in\Omega_i} \sum_m \left(mQ_{ij} - \mu_i \left(\frac{2^m - 1}{h_{ij}^k}\right)\right) \rho_{ij}^{(k,m)} + \mu_i P_i, \end{aligned}$$

where μ_i is the Lagrangian multiplier associated with (24). The corresponding dual function is

$$\begin{aligned} D_i(\mu_i) &= \sup_{\rho:k\in\Omega_i} \{L_i(\rho, \mu_i) : (22), (23)\} \\ &= \sup_{\rho:k\in\Omega_i} \{ \sum_{k\in\Omega} \sum_j \sum_m \left(mQ_{ij} - \mu_i \left(\frac{2^m - 1}{h_{ij}^k}\right)\right) \rho_{ij}^{(k,m)} \\ &:(22), (23)\} + \mu_i P_i. \end{aligned} \tag{26}$$

We evaluate $D_i(\mu_i)$, for a given μ_i, by decomposing it into subproblems of link and modulation rate assignment to each subcarrier $k \in \Omega_i$ as

$$\text{P7}: \quad \max_\rho \sum_j \sum_m \left(mQ_{ij} - \mu_i (\frac{2^m - 1}{h_{ij}^k})\right) \rho_{ij}^{(k,m)} \tag{27}$$

$$\text{s.t.} \quad \sum_j \sum_m \rho_{ij}^{(k,m)} = 1 \tag{28}$$

$$\rho_{ij}^{(k,m)} \in \{0,1\}, \ \forall m, j. \tag{29}$$

According to (28) and (29), subcarrier k can be assigned to only one outgoing link with one modulation rate. This subcarrier should be assigned to link ij_k with modulation rate m_k according to

$$(j_k, m_k) = \arg\max_{(j,m)} \left(mQ_{ij} - \mu_i \left(\frac{2^m - 1}{h_{ij}^k}\right)\right). \tag{30}$$

In other words, $\rho_{ij}^{(k,m)} = 1$ if $(j, m) = (j_k, m_k)$ for each $k \in \Omega_i$, otherwise is 0. Dual variable μ_i is obtained from the following dual problem

$$\text{P8}: \min_{\mu_i \geqslant 0} D_i(\mu_i). \tag{31}$$

Problem P8 is solved by subgradient method. Starting with an initial value μ_i^1, at each iteration τ with a given μ_i^τ, the optimal pair (j_k^τ, m_k^τ) for each subcarrier $k \in \Omega_i$ is obtained from (30), and then $\mu_i^{\tau+1}$ is updated as

$$\mu_i^{\tau+1} = \left[\mu_i^\tau - \sigma \left(P_i - \sum_{k\in\Omega_i} \frac{2^{m_k^\tau} - 1}{h_{ij_k^\tau}^k} \right) \right]^+, \tag{32}$$

where $(P_i - \sum\limits_{k\in\Omega_i} \frac{2^{m_k^\tau} - 1}{h_{ij_k^\tau}^k})$ is the subgradient of $D_i(\mu_i^\tau)$ with respect to μ_i^τ in (26), and σ is the step size. At time slot t, we initialize μ_i, with its optimal value achieved in time slot $t-1$ to speed up the convergence.

Since modulation rates, m_k's , are not continuous, equation (32) does not converge to a stable point; but converges to a stable boundary. This fact implies that the duality gap is not exactly zero in P6 due to the non-convexity of the feasible region [32]. To stop the iterations, we use the terminating condition

$$\left| P_i - \sum_{k\in\Omega_i} \frac{2^{m_k^\tau} - 1}{h_{ij_k^\tau}^k} \right| < \epsilon, \tag{33}$$

where ϵ is a small enough value. We investigate the impact of non-zero duality gap, on the aggregate utility in Section VI. The scheduling algorithm for node i, is summarized in Algorithm 3.

Algorithm 3 JSARA scheduling scheme at each node i

1: **Input**: $Q = \{Q_{ij}\}$, $H = \{h_{ij}^k\}$.
2: **Output**: $\rho = \{\rho_{ij}^{(k,m)}\}$.
3: set $\tau = 1$ and initialize μ_i^τ.
4: **while** (1) **do**
5:   **for all** $k \in \Omega_i$ **do**
6:     given μ_i^τ, obtain (j_k^τ, m_k^τ) from (30).
7:   **end for**
8:   **if** $|P_i - \sum_{k \in \Omega_i} \frac{(2^{m_k^\tau}-1)}{h_{ij_k^\tau}^k}| \geq \epsilon$ **then**
9:     update μ_i^τ by (32).
10:     $\tau = \tau + 1$, go to step 5.
11:   **else**
12:     break.
13:   **end if**
14: **end while**
15: **for all** $k \in \Omega_i$ **do**
16:   **if** $(j, m) = (j_k, m_k)$ **then**
17:     $\rho_{ij}^{(k,m)} = 1$.
18:   **else**
19:     $\rho_{ij}^{(k,m)} = 0$.
20:   **end if**
21: **end for**
22: return $\rho = \rho_{ij}^{(k,m)}$.

V. HEURISTIC ALGORITHM

In the greedy algorithm, more than one outgoing link can be active simultaneously at any node during each time slot. Therefore, the complexity at node i is $O(N_{\mu_i} \times |\Omega_i| \times N_i \times M)$, where N_{μ_i} is the number of iterations required for the convergence of (32), N_i is the number of outgoing links, and M is the number of modulation rates. To reduce the complexity and make the scheduling more tractable in real-time, we propose a heuristic algorithm where only one outgoing link can be active at any node during each time slot. Similar to the greedy algorithm, subcarrier set Ω is assigned to the network nodes according to Algorithm 2. Afterwards, at node i, we consider a TDMA scheme where all assigned subcarriers, $k \in \Omega_i$, are allocated to only one outgoing link during each time slot. Therefore, the complexity reduces to $O(N_i + N_G \times |\Omega_i|)$, where N_G is the number of iterations required for the convergence of bit loading algorithm.

Considering the objective function in (21), maximizing sum (rate $\times$ queue-length), the heuristic algorithm activates link ij^* at node i, exclusively, where

$$j^* = \arg\max_j \left(Q_{ij} \sum_{k \in \Omega_i} m_{ij}^k \right) \tag{34}$$

and $m_{ij}^k = \lfloor \log_2(1 + h_{ij}^k p_{ij}^k) \rfloor$. Moreover, we employ the bit loading algorithm proposed in [20] to obtain p_{ij}^k, $k \in \Omega_i$, the optimal allocated power to subcarrier k, if it is assigned to link ij. The heuristic algorithm is summarized in Algorithm 4.

Algorithm 4 Heuristic algorithm

1: **Input**: $Q = \{Q_{ij}\}$, $H = \{h_{ij}^k\}$.
2: **Output**: $\rho = \{\rho_{ij}^{(k,m)}\}$.
3: obtain Ω_i for all i using Algorithm 2.
4: **for all** i **do**
5:   $j^* = \arg\max_j (Q_{ij} \sum_{k \in \Omega_i} m_{ij}^k)$.
6:   assign all subcarriers in Ω_i to link ij^*.
7:   compute modulation rates using bit loading algorithm.
8: **end for**
9: return $\rho = \rho_{ij}^{(k,m)}$.

VI. PERFORMANCE EVALUATION

In this section, we evaluate the cooperation performance of the rate control mechanism and the scheduling schemes. We evaluate the performance when the greedy, the heuristic, and the distributed hierarchical scheduling (DHS) [24] algorithms are applied. DHS is a fair multi-hop scheduling which assigns the subcarriers to the nodes based on the channel gains averaged over all subcarriers, followed by the power allocation. We will compare the fairness of our scheme , i.e., proportional fairness[1], with that of the DHS, which employs a fair subcarrier assignment to the network nodes.

We consider downlink transmission in an OFDMA WMN, shown in Fig. 3, where the BS serves five flows, $s = 1, 2, ..., 5$, with corresponding destinations d_s. There are 128 subcarriers over one MHz frequency band, and the total transmission power of each transmitter is 10 Watts. The fading channel on each link ij is a 6-tap Rayleigh fading with 0.9 μs RMS delay spread. The channel exponential power delay profile is $g_{ij} e^{-(l-1)}$, where g_{ij} is the first path's average power gain, and l is the path index. We assume all links have the same channel fading gain $g_{ij} = 0\, dB$. Also, single-sided power spectral density of noise is unity. We perform the simulation for 1500 realizations of a fading channel, i.e., 1500 time slots.

[1] To address fairness, we consider logarithmic utility functions in P1, i.e., $U_s(r_{BS}^{(s)}) = \log(r_{BS}^{(s)})$, to maintain proportional fairness.

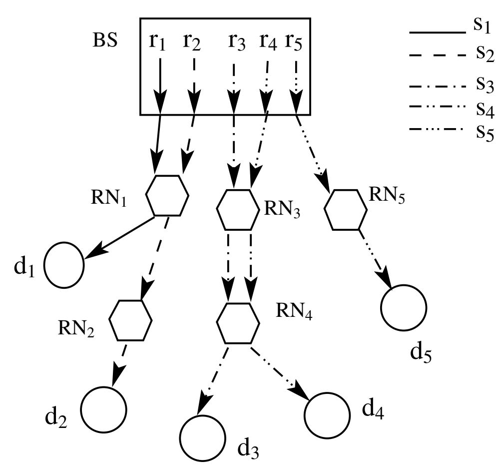


Fig. 3. OFDMA WMN architecture

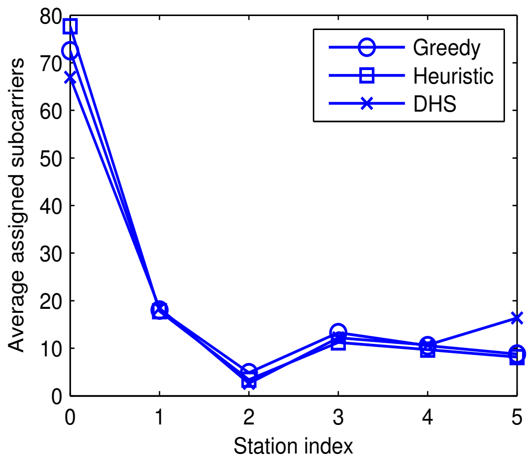


Fig. 4. Average assigned subcarriers to nodes

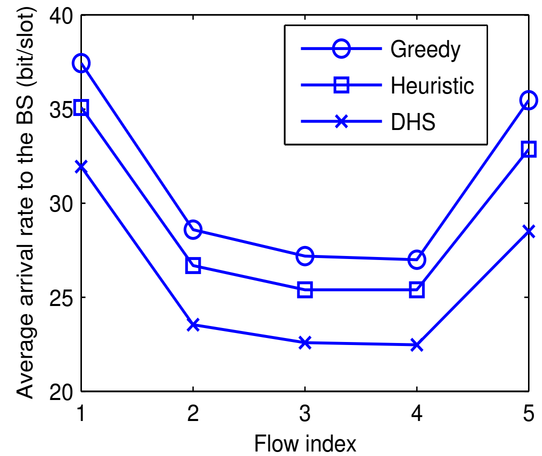
The average number of subcarriers allocated to each node by the examined scheduling schemes is shown in Fig. 4. Node 0 represents the BS, node 1 is RN_1,..., and so on. It is observed that the number of subcarriers assigned to each node is approximately proportional to the load, i.e., the number of flows, served by that node (see Fig. 3). Especially, the BS, which serves all the flows, has been assigned the largest number of subcarriers. The difference between the greedy and the heuristic algorithms in Fig. 4 results from different subcarrier and modulation rate assignments in individual nodes. Furthermore, the difference between our proposed schemes and DHS in Fig. 4 arises from the fact that we employ individual subcarriers' average channel gain in Algorithm 2, while DHS uses the average channel gains over all subcarriers. Therefore, we achieve more efficient subcarrier assignments because of more accurate channel state information.

The average arrival rates and the BS queue lengths of the flows are shown in Fig. 5a and Fig. 5b, respectively. According to these figures, as a queue length increases, the allocated arrival rate decreases, which conforms the rate control equation in (14), under the assumption of logarithmic utility function in this paper. Furthermore, the heuristic algorithm demonstrates comparable performance to that of the greedy algorithm (6% of performance loss in average). Flows 1 and 5 which have short path lengths, i.e., 2 hops, have higher arrival rates and smaller queue lengths. Also, flows 2, 3, and 4 which have the same path length, i.e., 3 hops, have approximately the same performance. Since DHS solution assigns subcarriers to the network nodes, based on the channel gains averaged over all subcarriers, both our proposed algorithms outperform this solution significantly, i.e., approximately by 18%.

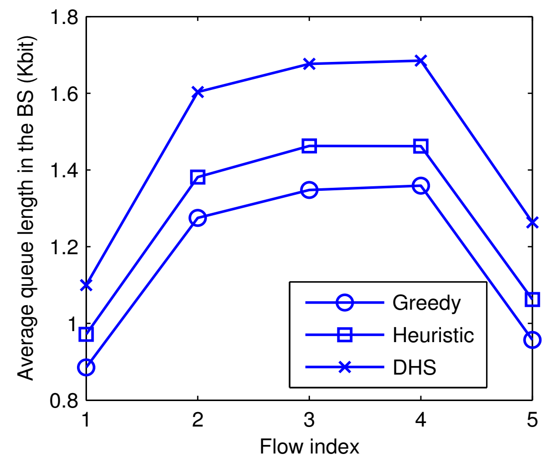
The average aggregate queue length and the average end-to-end delay, experienced by each flow, are shown in Fig. 6a and Fig. 6b, respectively. The aggregate queue length of each flow is defined as the sum of queue lengths of the nodes on its path. As expected, flows with smaller path lengths, i.e., flows 1 and 5, have smaller aggregate queue length and delay than flows 2, 3 and 4 with higher path lengths. In addition, flow 2 has smaller delay. This is due to the fact that flow 2 has only one shared link on its path, while flows 3 and 4 have 2 shared links on their paths. Considering Fig. 6 and the average arrival rates in Fig. 5a, we conclude that our observations on the network performance conforms Little's law [33]. In other words, the average queue length is equal to the average arrival rate multiplied by the average delay.

To compare the scheduling algorithms in terms of fairness, we compute Jain's fairness index [34] of the allocated arrival rates. As shown in Table 1, our proposed schemes have comparable performance to that of DHS which is considered as a fair scheduling scheme. The minor outperformance of our algorithms is because of the more accurate subcarrier assignments.

Finally, the aggregate utility, achieved over simulation time, when the rate control and the scheduling schemes are cooperating, is depicted in Fig. 7. To investigate the duality gap



(a)

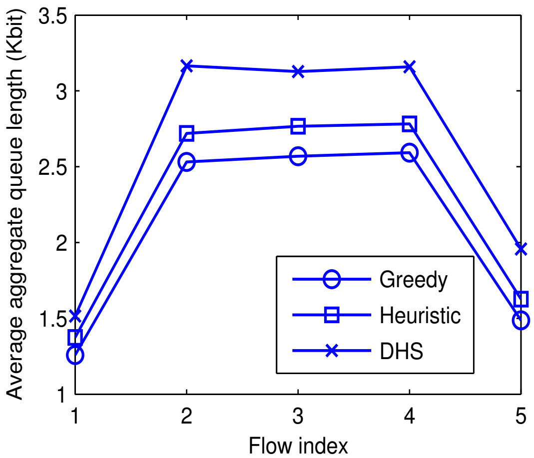


(b)

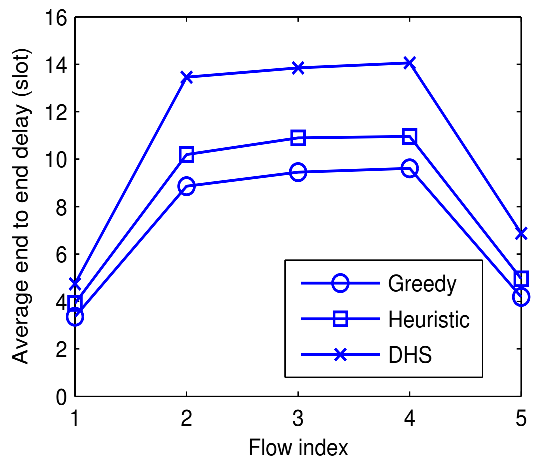
Fig. 5. (a) The average arrival rate of the flows (b) The average queue length in the BS

TABLE I
FAIRNESS INDEX OF ARRIVAL RATES

Algorithm	Greedy	Heuristic	DHS
Fairness	0.98	0.98	0.97



(a)



(b)

Fig. 6. (a) The average aggregate queue length of the flows (b) The average end-to-end delay of the flows

of the greedy algorithm, we compute the aggregate utility of the continuous solution of the scheduling problem. In the continuous solution, we relax the discrete rate allocation to subcarriers in order to obtain an upper bound on the aggregate utility. Despite the time-varying channels, the aggregate utility curves converge to a stable value after a while which depends on the step size value in (14). The duality gap is bounded within 0.25% of the aggregate utility. This result conforms to the ones in [23], [28], i.e., the duality gap approaches zero as the number of subcarriers increases. In addition, the heuristic algorithm demonstrates comparable performance to that of the greedy algorithm and outperforms DHS solution. This observation is consistent with the average arrival rates in Fig. 5a.

In summary, the cooperation between the rate control mechanism and the scheduling schemes improves the network performance in terms of aggregate utility.

VII. Conclusion

A cross-layer resource allocation problem for network performance optimization has been presented for OFDMA

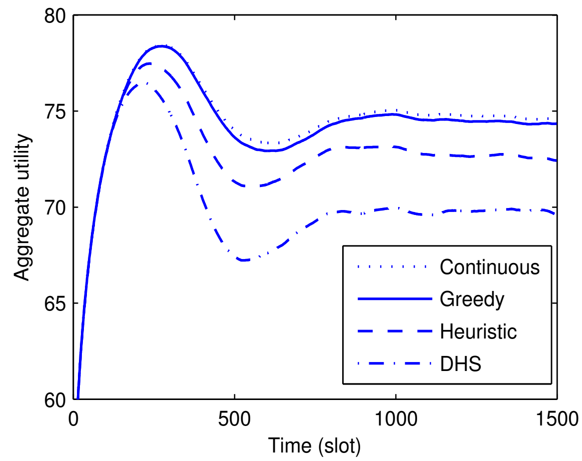


Fig. 7. Aggregate utility

wireless mesh networks. Solving the problem, by a dual decomposition, results in a rate control mechanism which controls the arrival rates at the BS, and a joint channel-aware and queue-aware scheduling scheme which determines the departure rates from the nodes. We have proposed a greedy and a heuristic algorithm for the joint channel-aware and queue-aware scheduling problem. Performance evaluation of our proposed scheduling schemes shows that the number of subcarriers assigned to each node is approximately proportional to the load, i.e., the number of flows, served by that node. Besides, when the scheduling cooperates with the rate control mechanism, the arrival rates to the BS are allocated based on the link rates provided by the scheduling scheme. This cooperation results in higher arrival rate and lower end-to-end delay of flows with smaller path lengths and thus improves the aggregate utility. Furthermore, the proposed heuristic algorithm demonstrated comparable performance to that of the greedy algorithm while having lower computational complexity. To enhance the capacity of the network, we aim to develop frequency reuse in our future works.

References

[1] F. Akyildiz and X. Wang, "A Survey on Wireless Mesh Networks, " *IEEE Communications Magazine*, vol. 43, no. 9, pp. 23-30, Sept. 2005.

[2] M. J. Lee and J. Zheng, "Emerging Standards for Wireless Mesh Technology, " *IEEE Wireless Communications*, vol. 13, no. 2, pp. 56-63, Apr. 2006.

[3] M. Salem, A. Adinoyi, H. Yanikomeroglu, and D. Falconer, "Opportunities and Challenges in OFDMA-Based Cellular Relay Networks: A Radio Resource Management Perspective," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 5, pp. 2496-2510, Jun. 2010.

[4] H. Jiang, W. Zhuang, and X. Shen, "Cross-layer Design for Resource Allocation in 3G Wireless Networks and Beyond," *IEEE Communications Magazine*, vol. 43, no. 12, pp. 120-126, Dec. 2005.

[5] L. Chen, S. Low, and J. Doyle, "Joint Congestion Control and Media Access Control Design for Ad Hoc Wireless Networks," *in Proc. IEEE INFOCOM*, Mar. 2005.

[6] J. Lee, M.Chiang, and Calderbank, "Jointly Optimal Congestion and Contention Control in Wireless Ad Hoc Networks " *IEEE Communication Letters*, vol. 10, Mar. 2006.

[7] X. Wang and K. Kar, "Cross-layer Rate Optimization for Proportional Fairness in Multihop Wireless Networks with Random Access," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1548-1559, Aug. 2006.

[8] L. Chen, S. Low, and M. Chiang, "Cross-layer Congestion Control, Routing and Scheduling Design in Ad Hoc Wireless Networks," *in Proc. IEEE INFOCOM*, 2006.

[9] T. ElBatt and A. Ephremides, "Joint Scheduling and Power Control for Wireless Ad Hoc Networks," *IEEE Transactions on Wireless Communications*, vol. 3, no. 1, pp. 74-85, Jan. 2004.

[10] Y. Li and A. Ephremides, "A Joint Scheduling, Power Control, and Routing Algorithm for Ad Hoc Wireless Networks," *Ad Hoc Networks*, vol. 5, no. 7, pp. 959-973, Sept. 2007.

[11] X. Lin and N. Shroff, "Joint Rate Control and Scheduling in Wireless Networks," *in Proc. IEEE CDC*, Dec. 2004.

[12] X. Lin and N. Shroff, "The Impact of Imperfect Scheduling on Cross-Layer Congestion Control in Wireless Networks," *IEEE/ACM Transactions on Networking*, vol. 14, no. 2, pp. 302-315, Apr. 2006.

[13] M. J. Neely, E. Modiano, and C. Li, "Fairness and Optimal Stochastic Control for Heterogeneous Networks," *in Proc. IEEE INFOCOM*, Mar. 2005.

[14] A. Eryilmaz and R. Srikant, "Joint Congestion Control, Routing and MAC for Stability and Fairness in Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1514-1524, Aug. 2006.

[15] A. L. Stolyar, "Maximizing Queueing Network Utility Subject to Stability: Greedy Primal-Dual Algorithm," *Queueing Systems*, vol. 50, pp. 401-457, 2005.

[16] F. Kelly, "Charging and Rate Control for Elastic Traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33-37, 1997.

[17] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, "Layering as Optimization Decomposition: A Mathematical Theory of Network Architectures," *Proceedings of IEEE*, vol. 95, no. 1, pp. 255-312, Jan. 2007.

[18] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu, "Impact of Interference on Multi-Hop Wireless Network Performance, " *Wireless Networks*, vol. 11, no. 4, pp. 471-487, July 2005.

[19] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive Resource Allocation in Multiuser OFDM Systems With Proportional Rate Constraints," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2726-2737, Nov. 2005.

[20] G. Song and Y. Li, "Cross-Layer Optimization for OFDM Wireless Networks-Part II: Algorithm Development," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 625-634, Mar. 2005.

[21] C. Mohanram and S. Bhashyam, "Joint Sub-carrier and Power Allocation in Channel-Aware Queue-Aware Scheduling for Multiuser OFDM," *IEEE Transactions on Wireless Communication*, vol. 6, no. 9, pp. 3208-3213, Sept. 2007.

[22] C. Y. Wong, R. S. Cheng, K. B. Lataief, and R. D. Murch, "Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 10, pp. 1747-1758, Oct 1999.

[23] K. Seong, M. Mohseni, and J. M. Cioffi, "Optimal Resource Allocation for OFDMA Downlink Systems," *in IEEE International Symposium on Information Theory*, July 2006.

[24] K. D. Lee and V. C. M. Leung, "Fair Allocation of Subcarrier and Power in an OFDMA Wireless Mesh Network," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 11, pp. 2051-2060, Nov. 2006.

[25] B. G. Kim and J. W. Lee, "Joint Opportunistic Subchannel and Power Scheduling for Relay-Based OFDMA Networks With Scheduling at Relay Stations," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 5, pp. 2138-2148, Jun. 2010.

[26] W. Wenyi and W. Renbiao, "Capacity Maximization for OFDM Two-Hop Relay System With Separate Power Constraints," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 9, pp. 4943-4954, Nov. 2009.

[27] M. Kaneko, P. Popovski, and K. Hayashi, "Throughput-Guaranteed Resource-Allocation Algorithms for Relay-Aided Cellular OFDMA System," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 1951-1964, May 2009.

[28] W. Yu and R. Lui, "Dual Methods for Nonconvex Spectrum Optimization of Multicarrier Systems," *IEEE Transactions on Communications*, vol. 54, no. 7, pp. 1310-1322, July 2006.

[29] D. Berteskas, *Nonlinear Programming*. Boston: Athena Scientific, 1999.

[30] L. Tassiulas, "Scheduling and Performance Limits of Networks with Constantly Varying Topology," *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 1067-1073, May 1997.

[31] B. Korte and J. Vygen, *Combinatorial Optimization: Theory and Algorithms*. New York: Springer-Verlag, 2002.

[32] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[33] J. King, *Computer and Communication Systems Performance Modeling*. New York: Prentice-Hall, 1990.

[34] R. Jain, *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation and Modeling*. New York: Wiley, 1991.

Mohammad Fathi received the M.Sc. degree in electrical engineering from Amirkabir University of Technology, Tehran, Iran in 2003. He was a lecturer at the Department of Electrical Engineering, University of Kurdistan, Sanandaj, Iran from 2003 to 2006.

Currently, he studies toward the Ph.D. degree at the Department of Electrical Engineering, Amirkabir University of Technology. Since February 2010, he has been a visiting Ph.D. student at the Communication Theory Laboratory, Royal Institute of Technology (KTH), Sweden. His research interests are in the area of wireless network resource allocation and optimization.

Hassan Taheri received his B.S. degree in electronics engineering from Amirkabir University of Technology, Tehran, Iran; M.S. and Ph.D. from University of Manchester Institute of Science and Technology in1975, 1978 and 1988 respectively. He is an associate professor in department of electrical engineering, Amirkabir university of Technology, Tehran, Iran. His research interests include Teletraffic Engineering and Quality of Service in fixed and mobile communication networks.

Mehri Mehrjoo (M'06) received the B.A.Sc. and the M.A.Sc. degrees from Ferdowsi University, Mashad, Iran, and Ph.D. from the University of Waterloo, Waterloo, Canada in 1993, 1996, and 2008, respectively.

From 2008 to 2009, she has been a post-doctoral fellow at the University of Waterloo. Currently, she is an assistant professor in the Department of Telecommunications, University of Sistan and Baluchestan, Zahedan, Iran. Her research interests are in the areas of resource allocation, performance analysis of wireless protocols, vehicular communication.