Cross Layer Optimization with Complete Fairness Constraints in OFDMA Relay Networks

Jianwei Wang1,2,3, Yuping Zhao3, Timo Korhonen4
Shenzhen Institute of Advanced Technology1
The Chinese University of HongKong2
Satellite and Wireless Communication Lab, Peking University, China3
Communications Engineering Lab, Helsinki University of Technology, Finland4

Abstract—Orthogonal frequency division multiple access (OFDMA) relay networks have been drawing much more attention in recent years. Cross layer optimization has been shown to be an efficient tool for resource allocation in OFDMA relay networks. However, the fairness issues for OFDMA relay networks have not been examined clearly in existing literature. In this paper, we considered a relay scenario with multiple sources, multiple relays and a single destination. A complete-fairness cross layer optimization framework is proposed, which considers fairness for both sources and relays. The simulation results show that the proposed algorithm outperforms the existing fairness algorithms significantly.

Index Terms—Cross Layer, OFDMA, Relay, Fairness

I. INTRODUCTION

In relay network, a set of nodes act as data forwarders and relay traffic for other nodes, therefore the network capacity can be improved. Currently, 802.16j Mobile Multihop Relay (MMR) working group is focused on integrating relay schemes into 802.16-based network with centralized or semidistributed resource allocation [1]. By introducing relay stations (RS), 802.16j tries to realize data communication with relay functionality of RS between MMR base stations (BS) and mobile stations (MS). Optimized link topology will also be established to lead an efficient utilization of network bandwidth [2]. In contrast to traditional signal amplifier, RS is capable of employing complex signal processing techniques and advanced communication schemes, which help to enhance both the network coverage and the network throughput.

The classical relay channel was first modeled as a class of three-terminal communication channels originally examined by L. Ozarow in [3]. From the view of capacity, T. Cover and his research group examined discrete memoryless additive white Gaussian noise (AWGN) relay channels and developed the lower bounds on the channel capacity [4]. In paper [5], J. laneman further summarized the achievable data rate via three structurally different random coding schemes. Based on the fundamental work of [6], Gastpar et al. [7] studied the relay traffic pattern and improved the capacity of relay networks. From the view of relay schemes, three different kind of cooperative relay protocols have been examined by previous literatures, including fixed, selection and incremental relaying. In fixed relaying networks, the relays either amplify their received signals or decode and retransmit the messages to the destinations [8]. Selective relaying introduces signal noise ratio (SNR) based cooperation action [9], while incremental relaying improves the spectral efficiency by exploiting limited feedback [10]. More recently, space-time codes are employed to enhance the cooperation ability in relay networks [11].

Reviewing previous literatures on relay networks, the majority is focused on single source-destination pair traffic scenario. However, more general cases in practical application scenarios with multiple sources, relays and destinations require are still lack of consideration. [12] provided capacity theorems for multiple sources – single relay scenario, and [13] studied multiple sources – single relay cases. More recently, [14] investigated the fairness issues in an orthogonal frequency division multiple-access (OFDMA) uplink scenario with multiple sources, multiple relays and a single destination. However, [14] only considered the relay fairness from a view of energy efficiency without fairness constraints on sources. In this paper, we discuss the OFDMA uplink performance with fairness constraints on both sources and relays.

The rest of this paper is organized as follows. Section II describes the system model. The problem with dynamic fairness constraints is formulated in Section III. In Section IV, the problem solution based on graph theory is discussed. The simulation evaluation is provided in Section V and Section VI offers the concluding remarks.

II. SYSTEM DESCRIPTION

We consider a multi-carrier broadband OFDMA relay network with multiple sources, multiple relays and one destination as shown in Fig. 1. Our system structure here models the same relay scenario discussed in [14]. A typical interpretation of the model is uplink of the 802.16 MMR network [1]: the destination can be viewed as BS, the relay nodes can be viewed as relay stations defined in 802.16j, and the source nodes can be viewed as user terminals. Another interpretation is clustered Ad Hoc networks where the cluster header is the destination and the relays are selected from general nodes. Therefore, our model in Fig. 1 provides a general analysis model.

For simplicity, we make the following assumptions:

1) All the wireless nodes are assumed to work in half-duplex mode, i.e. the wireless nodes can not transmit and receive at the same time.

978-1-4244-2324-8/08/$25.00 © 2008 IEEE.
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2008 proceedings.
2) A centralized resource allocation is employed and the channel state information is known to the decision maker.

3) A robust protocol is employed to gather information and broadcast optimization decisions.

4) The channels are characterized to be slow fading and remain constant during resource allocation process.

5) A two-slot fixed relay mode is employed. In particular, in the first time slot, the source nodes transmit and the rest of the nodes receive. In the second time slot, the relay nodes transmit and the destination receive.

The half-duplex assumption above helps to make the relay network to be internode-interference (IDI) free. In such a IDI relay network, our task is to select relays for the sources and assign the frequency resources (subcarriers in OFDMA systems). For simplicity, we also assume that the relay nodes use the same subcarrier to relay information received from the same subcarrier, i.e. information will be relayed on the same subcarrier. This assumption simplifies the relaying process of the relay nodes and can highly reduce the searching time for optimal solution. As OFDMA systems employ multiple subcarriers and experience frequency selective fading, one subcarrier for a certain source-relay pair might be a better choice than another. Thus, our target here is to pursue the frequency diversity in the multi-carrier system by searching for the optimal subcarrier assignment. In the following, we will discuss the optimal resource allocation while considering fairness issues among both the sources and relays.

III. PROBLEM FORMULATION

We assume that there are \( L \) sources and \( K \) relays in the network as shown in Fig. 1. The subcarrier number is set to be \( N \). Denote the source node set as \( S = \{ S_1, S_2, \ldots, S_L \} \) and the relay node set as \( R = \{ R_1, R_2, \ldots, R_K \} \). Indicate the destination node as \( D \). Further denote the channel gains on subcarrier \( n \) from \( S_l \) to \( R_k \), \( D \) to \( R_k \) and \( S_l \) to \( D \) as \( c_{kl}^n \), \( c_{k}^n \) and \( a_{kl}^n \) respectively. Then in the first time slot, we get:

\[
y_k^n = a_{kl}^n x_l^n + v_k^n ,
\]

(1)

where \( x_l^n \) is the signal transmitted from \( S_l \) on subcarrier \( n \); \( y_k^n \) and \( y_l^n \) are the received signals at \( R_k \) and \( D \) in the first time slot; \( \hat{x}_k^n \) is the relayed signal from \( R_k \) in the second time slot; \( y_2^n \) is the received signal at \( D \) in the second time slot. \( v_k^n \), \( v_l^n \) and \( v_2^n \) denote independent and identically distributed (i.i.d.) Gaussian noise with variance \((N_0/2)\).

Here, we consider two types of relay schemes: amplified-and-forward (AF) scheme as well as decode-and-forward (DF) scheme. In AF relay, the relay nodes only transmit the amplified version of received signals from the sources without decoding. That is:

\[
\hat{x}_k^n = \mu_k^n y_k^n = \mu_k^n (a_{kl}^n x_l^n + v_k^n) ,
\]

(4)

where \( \mu_k^n \) is the power scaler constrained by the transmission power of \( \hat{x}_k^n \). On the other hand, the relay nodes decode the received signals before forwarding. [5] has analyzed the outage behavior of both AF and DF schemes in fixed relay model. [14] has provided a clear expression of the achievable data rate between \( S_l \) and \( D \), relayed by \( R_k \) on subcarrier \( n \):

\[
r_{AF}^{nl} = \frac{1}{2} \log_2 \left( 1 + \frac{|c_{kl}^n|^2}{N_0} + \frac{|\mu_k^n|^2 |a_{kl}^n|^2 |b_k^n|^2}{(\mu_k^n)^2 |b_k^n|^2 + 1} N_0 \right) .
\]

(5)

\[
r_{DF}^{nl} = \frac{1}{2} \min \left\{ \log_2 \left( 1 + \frac{|a_{kl}^n|^2}{N_0} \right) , \log_2 \left( 1 + \frac{|c_{kl}^n|^2 + |a_{kl}^n|^2 |b_k^n|^2}{N_0} \right) \right\} .
\]

(6)

We introduce a binary assignment variable \( \rho_{kl}^n \), where \( \rho_{kl}^n = 1 \) indicates the assignment of subcarrier \( n \) to \( S_l \) and \( R_k \). Otherwise, \( \rho_{kl}^n = 0 \). From a view of throughput optimal, it is intuitive to formulate the optimal resource allocation problem based on our model as follows:

\[
\text{Problem 1 (Basic optimal resource allocation problem):}
\]

\[
\text{Max } \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{l=1}^{L} \rho_{kl}^n r_{kl}^n ,
\]

Subject to:

\[
\rho_{kl}^n \in \{0, 1\}, \forall l, k, n ,
\]

(7)

\[
\sum_{k=1}^{K} \sum_{l=1}^{L} \rho_{kl}^n = 1 .
\]

(8)

where Eq. (8) forbids reuse of a certain subcarrier for multiple source-relay pairs. For simplicity, we represent the achievable data rate as \( r_{kl}^n \) in Problem 1.

Problem 1 only offers a framework to obtain the maximum throughput of the relay network without any fairness constraints. In [14], a relay fairness constraint is added and the energy consumption for multiple relays is balanced.
However, the fairness among multiple sources is still out of consideration. Moreover, the fairness constraint in [14] is not fair enough from the view of energy consumption. In this paper, we reformulate the optimal resource allocation problem into a more general framework, considering both the relay fairness and sources fairness. First, we related the relay load to the remaining energy of each relay node. In another word, the relay traffic allowed for each relay node will depend on the energy remains. Second, we constrain the number of subcarriers allocated to each source node based on the QoS requirements. Then we have the following problem:

**Problem 2:**

Max \( \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{l=1}^{L} p_{kl}^n r_{kl}^n \),  

Subject to:

\[
\rho_{kl}^n \in \{0, 1\}, \quad \forall m, k, n , \quad (9)
\]

\[
\sum_{k=1}^{K} \sum_{l=1}^{L} \rho_{kl}^n = 1 , \quad (10)
\]

\[
\sum_{n=1}^{N} \sum_{k=1}^{K} \rho_{kl}^n \leq p_k , \quad (11)
\]

where \( p_k \) and \( q_l \) constrain the number of subcarriers allocated to relay node \( k \) and source node \( l \) respectively. Denote the remaining energy of relay node \( r_k \) as \( E_k \), we get \( p_k \):

\[
p_k = \left[ \frac{N}{\sum_{k=1}^{K} E_k} \right] , \quad (13)
\]

where \( \left\lceil x \right\rceil \) returns the ceiling integer of \( x \). Comparing to the fairness limits in [14], Eq. (13) relates \( p_k \) with the remaining energy of relay node \( R_k \) and balances the fairness among relays more reasonably. On the other hand, denote the queue length and delay time of source node \( S_l \) as \( w_l \) and \( d_l \), we get:

\[
q_l = \left[ \frac{N}{\sum_{l=1}^{L} w_l \exp(\alpha d_l)} \right] , \quad (14)
\]

where \( \alpha \) is a scaling parameter to make tradeoffs between the priorities of the queue length and delay time. From Eq. (14), we can see that \( q_l \) considers both \( w_l \) and \( d_l \), which reflects the relative traffic demand of \( S_l \) from the view of QoS requirements.

Problem 2 is shown to be a linear integer optimal problem which has been proved to be NP-complete [15]. In the next section, we will focus on the solutions to Problem 2 from the point of view of graph theory.

**IV. Solution and Analysis**

In this section, we reformulate the problem and present the solution based on graph theory.

**A. Graph Theory Basis**

A graph is a pair \( G = (N, A) \) of sets such that \( A \subseteq [N]^2 \). Thus, the elements of \( A \) are 2-element subsets of \( N \). The elements of \( N \) are the nodes (or vertices, or points), and the elements of \( A \) are its arcs (or edges). For each arc \( j \), we denote it as \( j(i, i') \), where \( i \) is the initial node and \( i' \) is the terminal node. In this paper, we assume that there are no parallel arcs with the same set of initial node and terminal node, i.e. no multiplicities.

Let us view a graph as network: its arcs carry some kinds of flows. A flow \( x \) is a function \( x : A \rightarrow \mathbb{R} \). The value \( x(j) \), called flux in arc \( j \), is interpreted in most applications as the quantity of materials flowing in the arc \( j \). In our model, we assign a positive value \( x \) to the edge \( j(i, i') \) to express that a flow of \( x \) units passes through \( j \) from \( i \) to \( i' \), or assign \(-x\) to \( j(i, i') \) to express that \( x \) units of flow pass through \( j \) the other way, from \( i' \) to \( i \). A network \( G \) and its associated flows can be represented numerically aided by node-arc incidence function of \( G \), which is defined by:

\[
e(i, j) = \begin{cases} +1, & \text{if } i \text{ is the initial node of the arc } j; \\ -1, & \text{if } i \text{ is the terminal node of the arc } j; \\ 0, & \text{otherwise}. \end{cases}
\]

**Definition 1 (Divergence):** The divergence \( y \) associated with flow \( x \) at node \( i \) is defined as:

\[
y(i) = \sum_{j \in A} e(i, j) x(j) ,
\]

i.e.

\[
y = Ex = \text{div } x .
\]

In general, a node \( i \) is said to be a source for the flow \( x \) if \( y(i) > 0 \) and a sink if \( y(i) < 0 \). If \( y(i) = 0 \), the flow is conserved at \( i \). Physical intuition suggests, and algebra confirms, that the total amount created at the sources always equals the total amount destroyed at the sinks, which is expressed as total divergence principle:

\[
\sum_{i \in N} y(i) = 0, \quad \text{for } y = \text{div } x
\]

A path \( P \) in a network \( G \) is a finite sequence \( i_0, j_1, i_1, ..., j_r, i_r \) \((r > 0)\), where each \( i_k \) is a node, \( j_k \) is an arc, and either \( j_k \sim (i_{k-1}, i_k) \) or \( j_k \sim (i_k, i_{k-1}) \). The arc \( j_k \) in \( P \) is said to be traversed positively or negatively according to whether \( j_k \sim (i_{k-1}, i_k) \) or \( j_k \sim (i_k, i_{k-1}) \). A path \( P \) may traverse an arc more than once, maybe sometimes positively and sometimes negatively. Then it is a path with multiplicities. An elementary path is a path without multiplicities which in fact uses no node more than once, except for the initial and terminal nodes. For an elementary path \( P \), the set of all arcs traversed positively is denoted by \( P^+ \), and the set of all arcs traversed negatively is denoted by \( P^- \). Arc-path incidences are accordingly defined by:

\[
e_P(j) = e(j, P) = \begin{cases} +1, & \text{if } j \in P^+; \\ -1, & \text{if } j \in P^-; \\ 0, & \text{otherwise}. \end{cases}
\]
B. Linear Optimal Distribution Problem

In the optimal distribution problem, the goal is to minimize a sum of costs given by convex functions associated with the arcs of the network. Suppose each arc \( j \) of a connected network \( G \) has a capacity interval \([c^-(j), c^+(j)]\) and each node \( i \) has a supply \( b(i) \), where \( b(N) = 0 \). Suppose the cost of the flux \( x(j) \) is given by a linear expression \( d(j)x(j) + z(j) \), where \( d(j) \) and \( z(j) \) are constants associated with the arc \( j \). Thus, we have the following linear optimal distribution problem:

**Problem 3 (Linear Optimal Distribution Problem):**

\[
\text{Minimize } \sum_{j \in A} [d(j)x(j) + z(j)] = d \cdot x + \text{const} \tag{16}
\]

subject to:

\[
c^-(j) \leq x(j) \leq c^+(j), \text{ for all } j \in A, \tag{17}
\]

\[
y(i) = b(i), \text{ for all } i \in N \text{ (where } y(i) = \text{div } x). \tag{18}
\]

The constants \( z(j) \) could be dropped from (16) without any real loss of cost. The basic idea concerns a feasible solution \( x \) can be improved by sending flux around a circuit \( P \) which is flow augmenting and also satisfies:

\[
d \cdot e_p = \sum_{j \in P^+} d(j) - \sum_{j \in P^-} d(j) < 0. \tag{19}
\]

The flow-augmenting property means that for \( t > 0 \) sufficiently small, at least the flow \( x' = x + te_p \) will still satisfy the capacity constraints. Meanwhile, \( \text{div } x' = \text{div } x = b \).

C. Problem Transformation

We represent Problem 2 by a directed graph as shown in Fig. 2. The node set \( N \) includes:

- Subcarrier source nodes \( S^n \) with \( b(S^n) = 1 \).
- Source aggregate node \( S_a \) with \( N \) input branches and \( L \) output branches, where \( b(S_a) = 0 \).
- Source nodes \( SS_l \) representing the \( L \) user terminals in Fig. 1, where \( b(SS_l) = 0 \).

- Virtual source nodes \( S_{ln} \) generated from \( SS_l \) with \( b(S_{ln}) = 0 \). The number of \( S_{ln} \) is \( L \times N \).
- Virtual relay nodes \( R_{kn} \) with \( b(R_{kn}) = 0 \). The number of \( R_{kn} \) is \( K \times N \).
- Virtual converged relay nodes \( rR_k \) and \( \tilde{r}R_k \) with \( b(rR_k) = 0 \) and \( b(\tilde{r}R_k) = 0 \).

- Final sink node \( D \) with \( b(D) = -N \).

All the nodes except for \( S^n \) and \( D \) have a divergence requirement \( b = 0 \) and the generated graph satisfies the total divergence principle in (15). As \( b(S^n) = 1 \), the flux in the arc \( j(S^n, S_a) \) is 1, for \( n = 1, 2, ..., N \). The capacity interval of arc \( j(S_a, SS_l) \) is \([0, q_l]\), while the capacity intervals of arc \( j(rR_k, \tilde{r}R_k) \) and \( j(\tilde{r}R_k, D) \) are both \([0, p_k]\). All the other arcs in the network have a capacity interval \([0, 1]\). We set the cost value of arc \( j(S_{ln}, R_{kn}) \) as \(-r_{kl}\) and all the other arcs as 0. See Fig. 2 for more details. Now, we have generated a graph for Problem 2 and we can employ various algorithms such as simplex method [15] to solve the linear optimal distribution problem with integer capacity intervals.

D. Discussion and Theorems

For Problem 3, we have the following theorem [15]:

**Theorem 1:** If Problem 3 has an optimal solution, and if \( c^-(j), c^+(j), \) and \( b(j) \) are all integral for all nodes \( i \) and arcs \( j \), then there is an optimal solution that is integral.

From Theorem 1, we can conclude that Problem 2 has an integer optimal solution as \( c^-(j), c^+(j), \) and \( b(j) \) are all integral. More over, for Problem 2, we have:

**Theorem 2:** The feasible solution set is not empty if and only if: \( \sum_{l=1}^{L} q_l \geq N \) and \( \sum_{k=1}^{K} p_k \geq N \).

**Proof:** Suppose \( \sum_{k=1}^{K} p_k < N \) and the feasible solution is not empty. Cut \( G \) as \([D, N \setminus D]\), then according to max flow min cut theorem, the max flow transported to \( D \) is \( \sum_{k=1}^{K} p_k < N \). The divergence requirement \( b(D) = N \) can not be satisfied. Thus, \( \sum_{k=1}^{K} p_k \geq N \) is required. Similar conclusion can be acquired for \( \sum_{l=1}^{L} q_l \geq N \).

Recall that \( p_k, q_l \) in (13) and (14) satisfy \( \sum_{k=1}^{K} p_k \geq N \) and \( \sum_{l=1}^{L} q_l \geq N \) respectively, thus Problem 2 has feasible
V. SIMULATION RESULTS

In this section, we examine the performance of the proposed algorithm. Three different relay algorithms are considered: 1) random scheme to pair relays and sources randomly; 2) partial fairness scheme with constraints $p_k = \lceil N/K \rceil$ [14]; 3) complete fairness scheme proposed in this paper which considers both source fairness and relay fairness. We employ AF relay in a 64-subcarrier OFDMA system with $L = 4$, $K = 3$ during the simulation.

Fig. 3 shows the average sum rate per subcarrier vs. SNR. From Fig. 3, we can see that the capacity of complete-fair algorithm is a bit smaller compared to partial-fair algorithm. The reason for that is the introduction of constraint (12) in complete-fair algorithm. On the other hand, we define $U = \sum_{i=1}^{L} (g_i - g_{av})^2$ to represent the throughput unfairness among $S_i$, where $g_i$ is the average throughput of $S_i$ and $g_{av}$ is the average throughput of all $S_i$. Fig. 4 shows the value of $U$ under various channel gain of node $S_4$ while the channel gains of all the other nodes remain the same. The partial-fair algorithm is shown to produce the biggest $U$ when the channel gain of node $S_4$ increases, where a node with best channel gain starves the others. Thus, the complete-fair algorithm significantly outperforms the other two algorithms from a view of fairness.

VI. CONCLUSION

In this paper, we proposed a complete-fairness cross layer optimization framework, which considers fairness issues for both sources and relays. A directed graph is generated to solve the optimal resource allocation problem based on graph theory. The simulation results show that the proposed algorithm leads to a more fairer allocation comparing to existing partial fairness algorithms, while keeps the capacity nondecreasing.

ACKNOWLEDGMENT

This work is supported by the grant from Ministry of Science and Technology, The People’s Republic of China (International science and technology Cooperation Projects 2006DFB73360) and supported by the Chinese National High-Tech 863 Project (No. 2006AA012278).

REFERENCES