Cross-layer Optimization for Uplink Transmission in OFDMA Cellular Networks with Fixed Relays

Lei You[‡†] Ping Wu[‡] Mei Song[†] Junde Song[†] Yong Zhang[†]
[‡]Signals & Systems, Dept. of Engineering Sciences, Uppsala University, Uppsala, Sweden
[†]PCN&CAD Center, Beijing University of Posts and Telecommunications, Beijing, P.R. China
youlei.feng@gmail.com Ping.Wu@angstrom.uu.se {songm, jdsong, zhangyong}@bupt.edu.cn

Abstract—**In this paper we consider cross-layer design for uplink transmission in an OFDMA-based cellular network with fixed relay stations (RSs) aimed to enhance performance. Since mobile stations (MSs) spend most of the power in the uplink to transmission, power efficiency resource allocation becomes very important to MSs. We develop a cross-layer optimization framework for two types of uplink flows (inelastic and elastic flows) which have different QoS requirements. For inelastic flows with fixed rate requirement we formulate the cross-layer optimization problem as the minimization of the sum transmission power of MSs under the constraints of flow conservation law, subcarrier assignment, relaying path selection and power allocation. For elastic flows with flexible service rate requirement, we consider the cross-layer tradeoff between uplink service rate and power consumption of MSs and pose the optimization problem as the maximization of a linear combination of utility (of service rates) and power consumption (of MSs). Different tradeoff can be achieved by varying the weighting parameters. Dual decomposition and subgradient method are used to solve the problems optimally with reduced computational complexity. Simulation results show that through the proposed cross-layer resource optimization framework and algorithms, significant benefits of deployment of multiple fixed relays in an OFDMA cellular network can be fully obtained in the sense of reducing power consumption, increasing service rate and saving energy in the uplink transmission of MSs.**

Index Terms—**Power efficiency, Cross-layer Optimization, Dual Decomposition, Relays, OFDMA Cellular Networks**

I. INTRODUCTION

Relay-based deployment has been viewed as one of the most promising architectures for next generation cellular networks [1], because it can reduce subscribers' power consumption and deployment cost of the infrastructure, expand coverage of cells, and also enhance system capacity and throughput in cellular networks. Therefore, IEEE 802.16 Working Group (WG) has created the Relay Task Group (RTG) 802.16j [2] to add relay functionality to IEEE standard 802.16 and develop appropriate procedure for relaying operation.

In this paper relay-based OFDMA cellular networks are considered. OFDMA has capability of exploiting the frequency selectivity enabled multiuser diversity by adaptive resource allocation. A deep faded subcarrier for one node may be favored by another node. Therefore, the multiuser diversity may be exploited in a multiuser OFDM system, if subcarrier assignment for each user and power allocation for each subcarrier are appropriately adapted to the channel condition. Extensive research has been done to investigate resource allocation in traditional cellular without relays [3] [4] [5] [10] [26].

In order to attain best performance, optimal operation of relay-enhanced OFDMA cellular network should be employed. This, however, is not an easy task as it involves so many different combinations of power and subcarrier allocations and path selections especially when AMC (Adaptive Modulation and Coding) is used and QoS requirement of different flows should be considered. In addition, since a Mobile Station (MS) is a power- and energy-limited device, power-efficiency resource management and rate control strategies for the uplink transmission are important to increase lifetime of MSs. Although extensive research on relay-based OFDMA cellular networks has been done, the problems related to all the above aspects are not optimally solved. For example, in [6], subcarrier and power allocations to maximize system capacity in OFDMA relay cellular networks are considered, and a suboptimal approach dividing the problem into two heuristic steps is adopted. In [7], the optimal source, relay and subcarrier allocation problem with the fairness constraint on relays is solved using graph theoretical approach. However, it assumed that fixed power is allocated to each subcarrier, which can not yield optimal power control.

Recently, dual decomposition method is proved to be a computationally efficient method to obtain optimal solution in the resource allocation of multicarrier systems [8] [9]. Related work using such method in OFDMA cellular system includes weighted sum rate maximization and weighted sum power minimization for downlink [10] in traditional cellular network, utility maximization in user cooperation cellular network [11] and sum-rate maximization for downlink in relay-based OFDMA system [12]. In [11] optimal resource allocation and relay strategies (AF and DF) selection for user cooperation cellular network, in which subscribers can cooperatively forward data for each other but using the same OFDMA subcarrier as source node, is presented under utility maximization framework [13] [14]. In [12], similar problem for joint subcarrier and power allocation as [6] is formulated and solved by making continuous relaxation and using dual decomposition method. However, their objective is to maximize sum-rate for the downlink and they do not consider the QoS requirement of flows from upper layer, which would result in unfairness in resource allocation. In addition, since MS spend little power on the transmission in downlink case (a very small amount of power consumption for reception and signal processing), power efficiency of MSs on downlink is not an issue considered.

In [16], average achievable rate and average power consumption on the uplink transmission of cooperative OFDMA cellular networks with relay nodes, which have little concern about their power consumption, are analytically examined. Though it focused on a cooperative relaying scheme, the analytical results showed that the benefits in increasing average transmission rate and reducing average power consumption of MSs from the deployment of multiple relay nodes are significant. However, it did not address how to get those benefits and how to achieve best tradeoff between those benefits through intelligent and optimal resource allocation.

In this paper, we develop cross-layer resource optimization framework for the power efficiency of MSs and flows with different QoS requirements in the uplink transmission in OFDMA cellular networks with dedicated relays, which are assumed to be fixed and have unlimited energy but have maximum transmission power limit. First, we consider inelastic flows with specific rate requirement (e.g., VoIP services), and minimize the sum power consumption of MSs by optimally assigning subcarriers on direct and relaying uplinks and allocating power of RSs and MSs to subcarriers. Then, we consider elastic flows with flexible service rate (e.g., best-effort and non-real-time service) and investigate the cross-layer tradeoff between maximizing sum-utility of uplink service rate of elastic flows and minimizing power consumption of MSs by fully utilizing the relay nodes and the resource available. Thanks to the time-sharing property of multicarrier systems analyzed and showed in [9], the problem has zero duality gap, which enable us to solve it almost optimally in its dual domain using dual decomposition approach and subgradient iteration algorithm with reduced the computational complexity.

The remainder of paper is organized as follows. The system model considered is described in Section Ⅱ. Then, the resource optimization and solution for inelastic flows and power minimization are discussed in Section Ⅲ, and resource optimization and solution for tradeoff between uplink service rate of elastic flows and power consumption of MSs are presented in Section Ⅳ. After simulation results are discussed in Section Ⅴ, the paper is concluded in Section Ⅵ.

II. SYSTEM MODEL

In this section, we will present a system model for uplink transmission in OFDMA cellular networks with two-hop fixed relays.

A. *Network Model and uplink Frame Structure*

We consider the uplink transmission in an OFDMA relay-enhanced cellular network as shown in Fig. 1. In each cell, there are a Base Station (BS) at the centre and K fixed Relay Stations (RSs) evenly located around the BS. In uplink transmission, an MS can send signals either directly to the BS (referred to Direct Transmission, **DT**) or indirectly to the BS with two-hop through the help of one of the K RSs (referred to Relaying Transmission, **RT**). In this case, there are $K+1$ possible paths for a MS to communicate with the BS for uplink transmission. OFDMA subcarriers can be allocated separately to links between those nodes. The network model described here can be used to model uplink transmission of an IEEE 802.16j relay-based network [2].

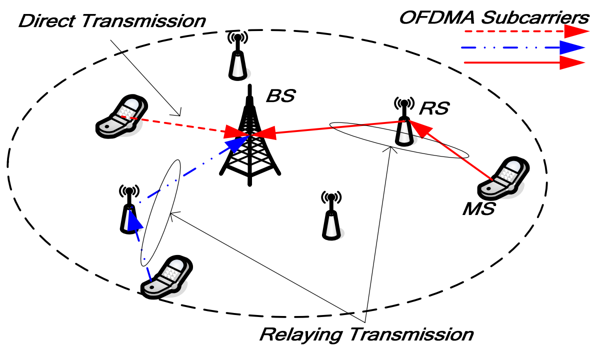

Fig.1 Uplink transmission model in OFDMA cellular network with two-hop fixed relays

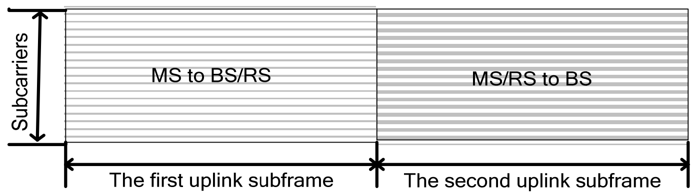

Fig.2 Frame structure for uplink transmission

We assume half-duplex operation of RSs on each subcarrier due to radio limitation. To guarantee proper transmission, we design a special uplink transmission frame with two-subframe as shown in Fig. 2. Each uplink frame includes two subframes. In the first subframe, MSs transmit the data to RSs or directly to BS. In the second subframe, RSs transmit the data they received from MSs in the first subframe to BS, while it is also possible for MSs to transmit directly to BS. Each subframe may use several OFDMA subcarriers in the frequency domain. To avoid interference, we impose that an OFDMA subcarrier can only be assigned to one of the uplink links MS-RS, MS-BS and RS-BS in any uplink subframe. We assume the lengths of the two subframes are the same. Such an uplink frame structure is slightly different from the one proposed in IEEE 802.16j MMR network [2], in which all MSs keep silent and only RSs are allowed to transmit in the second subframe. Our uplink frame structure can enable resource allocation more flexible.

We assume that wireless channels between nodes in the cellular are frequency-selective fading channels. OFDM technology divides the whole channel into many subcarriers so that each subcarrier experiences frequency-flat fading. We assume a slow-fading environment so that the channel remains unchanged during the resource allocation period. Full channel state information **(CSI)** is known to the BS which makes allocation decision in a centralized fashion and informs all

RSs and MSs of the results of the resource allocation through a certain reliable control channel. A similar network model is also considered in [6] and [12], and it is however focused on the downlink, while we consider the uplink case.

B. Power Allocation and Transmission Rate on Subcarriers in Physical Layer

Assume that there are K RSs labeled $\{1,...,k,...,K\}$ and M MSs randomly distributed in the cell, labeled $\{1,...,m,...,M\}$. The overall bandwidth B is divided into N OFDM subcarriers, labeled $\{1,...,n,...,N\}$. The channel coefficients of subcarrier n on the links MS m-to-BS, RS k-to-BS and MS m-to-RS k are $\gamma_{m,BS}^{n}$, $\gamma_{k,BS}^{n}$ and $\gamma_{m,k}^{n}$ respectively, the magnitudes of which follow a Rayleigh distribution. Consequently, the channel power gains $|\gamma_{m,BS}^{n}|^2$, $|\gamma_{k,BS}^{n}|^2$ and $|\gamma_{m,k}^{n}|^2$ follow the exponential distribution.

Let $p_m^{1,n}$ denote the power that MS m spends on subcarrier n during the first subframe. Let $p_m^{2,n}$ and $p_{k,BS}^{n}$ denote the power that MS m and RS k spend on subcarrier n during the second subframe, respectively. Note that an RS only spends power during the second subframe.

With the above parameters defined, we have the following transmission rate formulas for the subcarrier n

1) on link MS m-to-BS in the first uplink subframe

$$R_{m,BS}^{1,n} = W\log_2(1+p_m^{1,n}|\gamma_{m,BS}^{n}|^2/\Gamma W N_0)$$

2) on link MS m-to-RS k in the first uplink subframe

$$R_{m,k}^{1,n} = W\log_2(1+p_m^{1,n}|\gamma_{m,k}^{n}|^2/\Gamma W N_0)$$

3) on link MS m-to-BS in the second uplink subframe

$$R_{m,BS}^{2,n} = W\log_2(1+p_m^{2,n}|\gamma_{m,BS}^{n}|^2/\Gamma W N_0)$$

4) on link RS k-to-BS in the second uplink subframe

$$R_{k,BS}^{2,n} = W\log_2(1+p_{k,BS}^{n}|\gamma_{k,BS}^{n}|^2/\Gamma W N_0)$$

where $W = B/N$ is the bandwidth of each subcarrier, Γ is the SNR gap related to a targeted bit-error-rate, and N_0 is Additive White Gaussian Noise (AWGN) power spectral density, which is assumed to be the same for all the receiver nodes and subcarriers. In this case, given channel gains in the current frame, different power allocation may result in different transmission rates for subcarriers.

C. Subcarrier Allocation and Link Layer Rate

For the subcarrier allocation, we introduce binary indicators $d_{m,k}^{1,n}$, $d_m^{2,n}$ and $d_k^{2,n}$, which are explained below. Let $d_{m,k}^{1,n}=1$ represent that subcarrier n is allocated to the link MS m-to-RS k or MS m-to-BS (when k=0) in the first subframe, and $d_{m,k}^{1,n}=0$ otherwise; let $d_m^{2,n}=1$ represent that subcarrier n is allocated to the link MS m-to-BS in the second subframe, and $d_m^{2,n}=0$ otherwise; let $d_k^{2,n}=1$ represent that subcarrier n is allocated to the link RS k -to-BS in the second subframe, and $d_k^{2,n}=0$ otherwise. Those binary indicators must satisfy:

$$\sum_{k=0}^{K}\sum_{m=1}^{M} d_{m,k}^{1,n} = 1,\ d_{m,k}^{1,n} \in \{0,1\}, \forall n = 1,...,n,...,N \qquad (1)$$

$$\sum_{k=1}^{K} d_k^{2,n} + \sum_{m=1}^{M} d_m^{2,n} = 1,\quad d_k^{2,n}, d_m^{2,n} \in \{0,1\}, \forall n = 1,...,n,...,N \qquad (2)$$

where (1) means that for the first uplink subframe any subcarrier can only be assigned to one of the links MS-RS and MS-BS, while (2) states that for the second uplink subframe any subcarrier can only be assigned to one of the links RS-BS and MS-BS.

With the above assumptions and conditions, we can formulate the aggregate rates on links MS m-to-BS and MS m-to-RS k in the first subframe, and MS m-to-BS and RS k-to-BS in the second subframe, respectively, as follow:

$$T_{m,BS}^{1} = \sum_{n=1}^{N} d_{m,0}^{1,n} R_{m,BS}^{1,n},\ \forall m = 1,...,M$$

$$T_{m,k}^{1} = \sum_{n=1}^{N} d_{m,k}^{1,n} R_{m,k}^{1,n},\ \forall k = 1,...,K;\ \forall m = 1,...,M$$

$$T_{m,BS}^{2} = \sum_{n=1}^{N} d_{m}^{2,n} R_{m,BS}^{2,n},\ \forall m = 1,...,M$$

$$T_{k,BS}^{2} = \sum_{n=1}^{N} d_{k}^{2,n} R_{k,BS}^{2,n},\ \forall k = 1,...,K$$

Let $\boldsymbol{d}$ denote the vector of the binary indicators for a specific subcarrier allocation policy. Different subcarrier allocation policies result in different Link Layer Rates.

D. Flow Conservation Constraints for MSs and RSs.

Here we consider the cross-layer optimization for two types of uplink flows from MSs to BS: inelastic flow and elastic flow, respectively in the following sections. For inelastic flow, e.g. voice services, fixed service rate is usually required. Here, we assume that we have reliable coding and perfect admission control. Thus we can ignore other requirements such as Bit Error Rate (BER) and delay. For elastic flows such as non-real-time and best-effort services, which have no specific service rate requirements, some rate control schemes (e.g., TCP) are usually used to avoid network congestion and attain fairness.

We denote S_m (bits/frame) the total service rate of the uplink (inelastic or elastic) flows form MS m to the BS. For each MS m, the allocated aggregate uplink transmission rate in the two uplink subframes must be greater than or equal to S_m, and thus we have the following constraints,

$$S_m \le T_{m,BS}^{1} + \sum_{k=1}^{K} T_{m,k}^{1} + T_{m,BS}^{2},\ \forall m = 1,...,M \qquad (3)$$

For any RS k, the aggregate rate received in the first subframe must be less than or equal to the uplink rate of the link between the RS k and the BS in the second subframe. Thus we have

$$\sum_{m=1}^{M} T_{m,k}^{1} \le T_{k,BS}^{2},\ \forall k = 1,...,K \qquad (4)$$

III. RESOURCE OPTIMIZATION FOR INELASTIC FLOWS AND POWER EFFICIENCY OF MSS

In this section, we consider uplink inelastic flows with fixed required service rate.

In the cellular networks deployed with special fixed RSs for the purpose of performance enhancement, each RS has maximum power limitation although we have no concern about the energy expenditure of RSs. Letting $P_k^{\max}$ be the maximum power of RS k, we have the following constraints

$$\sum_{n=1}^{N} p_{k,BS}^{n} \le P_k^{\max},\ \ \forall k = 1,...,K \quad (5)$$

However, for MSs we need to concern more about their power and energy consumption. It is well-known that to attain the same transmission rate, transmission power increase exponentially with the distance to the receiver. The deployment of relay nodes can significantly reduce transmission power of the MSs far away from the BS in the cell for the same uplink rate requirement [1]. In order to achieve power efficiency transmission of inelastic flows under the network model assumption described in Section Ⅱ.A, whether an MS chooses relaying or not depends not only on its location but also on the service rate required by its inelastic flow. This will be shown below in a simple example.

A. A Simple Scenario with a single carrier

We consider a scenario in which a cell with radius 2000 m consists of one BS, one RS and one MS as showed in Fig. 3, and the distance between the RS and the BS is 1400 m. We assume that only one carrier is available, the channel gain only includes a large scale path loss component with path loss exponent of 4, the bandwidth of the channel is $W = 100\text{kHz}$, $\Gamma = 1$, and $N_0 = -174\text{dBm}$.

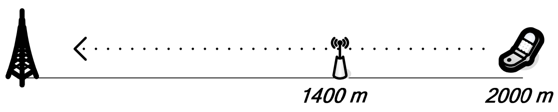

Fig.3 A simple scenario

From the network model described in SectionⅡ.A, the MS can use DT and transmit data directly to BS in both uplink subframes, or use RT and transmit data to the RS in the first subframe and then the RS forwards the data to BS in the second subframe. For the RT case, we assume that RS has sufficient power to transmit to the BS all the data coming from MS. Thus, MS spends its power in both subframes for DT, while just spends power in the first subframe for RT.

We calculate the power consumption of the MS using the relations given in Section II.B. The results are showed in Fig. 4 when the required data rate is 0.5Mbit/s and Fig. 5 when the required data rate is 1Mbit/s. From Fig. 4 and 5, we can see that: 1) the deployment of RSs can significantly reduce the level of uplink transmission power of the MS when it is far away from the BS, especially when it is located near the boundary of the cell; 2) to save power as much as possible, how an MS chooses a transmission scheme (DT or RT) depends not only on its position in the cell but also on the service rate required by the inelastic flow. For example, when the MS is located at a distance between about 800 m-950m to the BS, it should choose RT scheme when the required service rate is 0.5Mbit/s (see Fig. 4), but DT when the required service rate is 1Mbit/s (see Fig. 5).

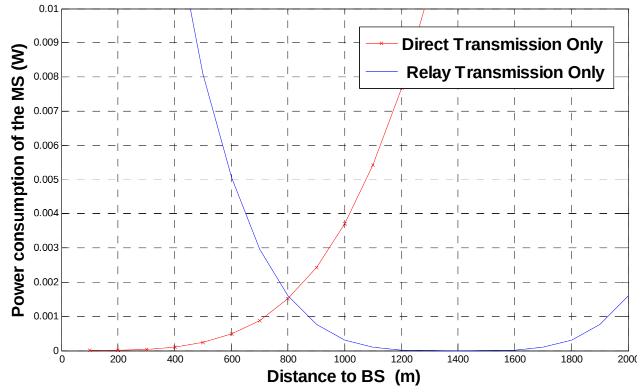

Fig.4 Power consumption of the MS at different locations and with different uplink transmission schemes (DT or RT) when the required service rate of the flow in the MS is 0.5Mbit/s

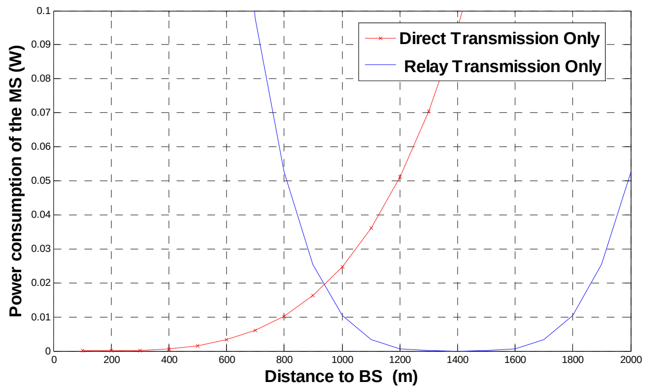

Fig.5 Power consumption of the MS at different locations and with different uplink transmission schemes (DT or RT) when the required service rate of the flow in the MS is 1Mbit/s. (Note that the vertical scale is ten times of that in Fig. 4)

This simple example shows that whether an MS chooses relaying or not depends on its location and the service rate required by its inelastic flow. For a more realistic scenario involving multi-carriers with frequency-selective fading channel, multiple RSs and MSs, optimal operation of the whole network to minimize power consumption of all the MSs will become much more complicated because it also depends on the channel gain of subcarriers between different nodes, power allocation as well as different rate requirements from all MSs. In addition, when the rate requirement is high, an RS may reach its maximum power, which will limit its capacity for relaying. In the following sub-section, we will tackle such a cross-layer resource allocation problem using nonlinear optimization techniques.

B. *Primal Problem Formulation, Dual Problem and Subgradient Method*

Form the above analysis, our objective is to find the optimal resource allocation solution to minimize the sum of power consumption of all the MSs in the two uplink subframes while the rate requirements from all the inelastic flows are satisfied under the constraints described in the system model. The optimization problem can be formulated as follows:

$$\text{Minimize } \sum_{m=1}^{M}(\sum_{t=1}^{2}\sum_{n=1}^{N} p_m^{t,n}) \qquad \textbf{(P1)}$$

Subject to (1) (2) (3) (4) and (5)

Note that the service rates $S_m, \forall m$ in constraints (3) of **P1** are known and fixed for all MSs. The optimization problem P1 is a NP-hard combination optimization problem with non-linear constraints [17], and finding its optimal solution involves an exhaustive search over all possible transmission schemes, RSs selection, subcarrier assignment policies and power allocations in the two uplink subframes, which is difficult to determine within a designated time, especially when the number of subcarriers is large because the dimension of the set of potential subcarrier assignment policies $\mathbf{d}$ increases exponentially with the number of subcarriers [10] [17].

It can be observed that (3), (4) and (5) are the coupling constraints of problem **P1** between two subframes and among different subcarriers. By relaxing these coupling constraints using the Lagrange multiplier technique, the problem can be decoupled into several subproblems that can be solved with low computational complexity given the Lagrange multipliers [8] [9] [15], and the optimal solution of the primal problem can be recovered optimally using gradient/subgradient method in its dual domain if strong duality holds [18].

However, due to binary variables in constraints (1) and (2), **P1** is a mixed integer programming problem and not a typical convex programming problem, thus strong duality may not hold. We can relax this integer constraint to a continuous one (i.e., let $d_{m,k}^{1,n}$, $d_m^{2,n}$ and $d_k^{2,n} \in [0,1]$) which corresponds to permitting time sharing of subcarrier allocation policies, and can change the problem to a convex one so that strong duality holds. This method needs to implement the resulting solution in multiple frames as in the TDMA-based stationary networks [5] [19] [20], which is impractical in the case of mobile cellular communication where the OFDMA channel varies from frame to frame [5].

Recently, it was discovered that the time-sharing condition under which the duality gap is zero is always satisfied in OFDM systems in the limit as the number of subcarriers goes to infinity as analyzed and proved in [9]. The reason is, roughly speaking, that in practical OFDM systems with a large number of subcarriers, channel conditions in adjacent subcarriers are often similar. Then, time sharing of each subcarrier may be approximately implemented with frequency sharing of these adjacent subcarriers [9] [10] [11]. Thanks to this result, we argue that the duality gap of **P1** is approximately zero and the dual method can still be used to solve **P1** optimally.

By introducing three vectors of Lagrange multipliers, $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_M]^T, \boldsymbol{\mu} = [\mu_1, ..., \mu_k]^T$ and $\boldsymbol{\varepsilon} = [\varepsilon_1, ..., \varepsilon_k]^T$ to relax coupling constraints (3), (4) and (5), the corresponding partial Lagrangian can be written as follows

$$\begin{aligned} & L_1(\mathbf{P_{MS}}, \mathbf{P_{RS}}, \mathbf{d}; \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\varepsilon}) \\ & = \sum_{m=1}^{M}(\sum_{t=1}^{2}\sum_{n=1}^{N} p_m^{t,n}) + \sum_{m=1}^{M}\lambda_m(S_m - T_{m,BS}^1 - \sum_{k=1}^{K} T_{m,k}^1 - T_{m,BS}^2) \\ & \quad + \sum_{k=1}^{K}\mu_k(\sum_{k=1}^{M} T_{m,k}^1 - T_{k,BS}^2) + \sum_{k=1}^{K}\varepsilon_k(\sum_{n=1}^{N} p_{k,BS}^n - P_k^{\max}) \end{aligned} \qquad (6)$$

where $\mathbf{P_{MS}}$ and $\mathbf{P_{RS}}$ are the vectors of power allocation for all MSs and RSs, respectively. Then, with this Langrangian we can define the dual objective function as

$$D(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\varepsilon}) = \begin{cases} \min\limits_{\mathbf{P_{MS}}, \mathbf{P_{RS}}, \mathbf{d}} L_1(\mathbf{P_{MS}}, \mathbf{P_{RS}}, \mathbf{d}; \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\varepsilon}) \\ s.t. \quad (1)\,(2) \end{cases} \qquad (7)$$

Thus the dual problem can be given as:

$$\begin{aligned} & \max\ D(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\varepsilon}) \\ & s.t.\ \boldsymbol{\lambda} \ge 0, \boldsymbol{\mu} \ge 0, \boldsymbol{\varepsilon} \ge 0 \end{aligned} \qquad \textbf{(D1)}$$

where Lagrange multipliers $\boldsymbol{\lambda}, \boldsymbol{\mu}$ and $\boldsymbol{\varepsilon}$ become the dual variables in dual problem **D1**. Due to the zero duality gap, i.e., strong duality holds, the solution of primal problem **P1** can be recovered by solving its dual problem **D1**. The most important advantage of solving the primal problem in its dual domain is the decomposability of the dual function, with which we can decouple the coupling constraints, decompose the problem into several subproblems and solve them separately with low complexity [15]. As showed in the next subsection, we will show that **D1** can be divided into separate per-subcarrier subprobms for each uplink subframe, which can significantly reduce the complexity of the problem.

Since the dual objective function $D(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\varepsilon})$ is not differentiable function, we can solve the dual problem using the subgradient method [21] [22]. To find subgradient using the definition in [21] [22], we first convert **D1** into the following an equivalent convex optimization problem with convex objective function.

$$\begin{aligned} & \min -D(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\varepsilon}) \\ & s.t.\ \boldsymbol{\lambda} \ge 0, \boldsymbol{\mu} \ge 0, \boldsymbol{\varepsilon} \ge 0 \end{aligned} \qquad \textbf{(D1')}$$

Then we can find a subgradient of the convex objective function in **D1'** by definition with the following Lemma.

Lemma 1: Considering the convex optimization problem **D1'**, and assuming that $T_{m,BS}^1{}^*, T_{m,k}^1{}^*, T_{m,BS}^2{}^*, T_{k,BS}^2{}^*$ and $p_{k,BS}^n{}^*$ are the optimal solution of minimization in the function (7) for given $\boldsymbol{\lambda}, \boldsymbol{\mu}$ and $\boldsymbol{\varepsilon}$, then

$$g_m(\lambda_m) = T_{m,BS}^1{}^* + \sum_{k=1}^{K} T_{m,k}^1{}^* + T_{m,BS}^2{}^* - S_m ,$$

$$h_k(\mu_k) = T_{k,BS}^2{}^* - \sum_{k=1}^{M} T_{m,k}^1{}^* \quad , \quad \text{and}$$

$$f_k(\varepsilon_k) = P_k^{\max} - \sum_{n=1}^{N} p_{k,BS}^n{}^* \quad ,$$

are the subgradients of $-D(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\varepsilon})$ at λ_m, μ_k and ε_k , respectively.

The proof of Lemma 1 is presented in the Appendix at the end of the paper.

Subgradient update algorithm [22] for solving **D1'** (and also

D1) is stated as follows

Given the optimal solution $T_{m,BS}^{1}{}^*, T_{m,k}^{1}{}^*, T_{m,BS}^{2}{}^*, T_{k,BS}^{2}{}^*$ and $p_{k,BS}^{n}{}^*$ in the current iteration i, the algorithm updates dual variable in the following manner

$$\begin{cases} \lambda_m(i+1) = \max(0, \lambda_m(i) - \sigma(i) g_m(\lambda_m(i))), \forall m \\ \mu_k(i+1) = \max(0, \mu_k(i) - \phi(i) h_k(\mu_k(i))), \forall k \\ \varepsilon_k(i+1) = \max(0, \varepsilon_k(i) - \theta(i) f_k(\varepsilon_k(i))), \forall k \end{cases} \quad (8)$$

where $\sigma(i)$, $\phi(i)$ and $\theta(i)$ are the step-sizes for the update of Lagrange multipliers $\boldsymbol{\lambda}, \boldsymbol{\mu}$ and $\boldsymbol{\varepsilon}$, respectively, in iteration i, until the algorithm converges.

According to [22], subgradient method is guaranteed to converge to the optimum if step-sizes $\sigma(i)$, $\phi(i)$ and $\theta(i)$ are designed appropriately as follows

Theorem 1: Dual variables $\boldsymbol{\lambda}, \boldsymbol{\mu}$ and $\boldsymbol{\varepsilon}$ converge to the optimal dual solutions if the positive scalar step-sizes $\sigma(i)$, $\phi(i)$ and $\theta(i)$ are chosen such that

$$\lim_{i\to\infty}\sigma(i) = 0, \sum_{i=1}^{\infty}\sigma(i) = \infty; \lim_{i\to\infty}\phi(i) = 0, \sum_{i=1}^{\infty}\phi(i) = \infty;$$

$$\lim_{i\to\infty}\theta(i) = 0, \sum_{i=1}^{\infty}\theta(i) = \infty$$

Remarks: Since strong duality holds, the corresponding primal variables ($\mathbf{P}_{MS}^*, \mathbf{P}_{RS}^*, \mathbf{d}^*$) are globally optimal variables of primal problem **P1** for optimal dual variables ($\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*, \boldsymbol{\varepsilon}^*$).

C. Dual Decomposition and Subproblems Solution

Each step of subgradient iteration algorithm requires the optimal solution of minimization in the dual objective function (7) which can be divided into the following two subprblems (9) and (10):

$$D(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\varepsilon}) = D_1(\boldsymbol{\lambda},\boldsymbol{\mu}) + D_2(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\varepsilon}) - \sum_{k=1}^{K}\varepsilon_k P_k^{\max}$$

where,

$$D_1(\boldsymbol{\lambda},\boldsymbol{\mu}) = \min_{\mathbf{P}_{MS}^1, \sum_{k=0}^{K}\sum_{m=1}^{M} d_{m,k}^{1,n}=1} \left(\sum_{m=1}^{M}\sum_{n=1}^{N} p_m^{1,n} - \sum_{m=1}^{M}\lambda_m\left(T_{m,BS}^1 + \sum_{k=1}^{K} T_{m,k}^1 \right) + \sum_{k=1}^{K}\mu_k \sum_{m=1}^{M} T_{m,k}^1 \right) \quad (9)$$

$$D_2(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\varepsilon}) = \min_{\mathbf{P}_{MS}^2, \mathbf{P}_{RS}, \sum_{k=1}^{K} d_k^{2,n} + \sum_{m=1}^{M} d_m^{2,n}=1} \left(\sum_{m=1}^{M}\sum_{n=1}^{N} p_m^{2,n} - \sum_{m=1}^{M}\lambda_m T_{m,BS}^2 - \sum_{k=1}^{K}\mu_k T_{k,BS}^2 + \sum_{k=1}^{K}\varepsilon_k \sum_{n=1}^{N} p_{k,BS}^n \right) \quad (10)$$

Given the dual variables, $D_1(\boldsymbol{\lambda},\boldsymbol{\mu})$ and $D_2(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\varepsilon})$ are to determine resource allocation in the first and second uplink subframes, respectively.

$D_1(\boldsymbol{\lambda},\boldsymbol{\mu})$ can be further written as

$$D_1(\boldsymbol{\lambda},\boldsymbol{\mu}) = \sum_{n=1}^{N}\left(\min_{\mathbf{P}_m^{1,n}, \sum_{k=0}^{K}\sum_{m=1}^{M} d_{m,k}^{1,n}=1} \sum_{m=1}^{M}\sum_{k=0}^{K}[p_m^{1,n} - (\lambda_m - \mu_k) d_{m,k}^{1,n} R_{m,k}^{1,n}] \right)$$

$$= \sum_{n=1}^{N}\left(D_1^n(\lambda_m, \mu_k) \right)$$

where $\mu_0 = 0$ and $R_{m,0}^{1,n} = R_{m,BS}^{1,n}$, and then decomposed to N subproblems as follows,

$$D_1^n(\lambda_m, \mu_k) = \min_{P_m^{1,n}, \sum_{k=0}^{K}\sum_{m=1}^{M} d_{m,k}^{1,n}=1} \sum_{m=1}^{M}\sum_{k=0}^{K} d_{m,k}^{1,n}[p_m^{1,n} - (\lambda_m - \mu_k) R_{m,k}^{1,n}], \quad \forall 1,...,n,...,N \quad (11)$$

Each subproblem in (11) is to determine the assignment of one subcarrier to one of the links between MSs and RSs/BS and also the amount of power of the MSs spent on it. The constraints in those N subproblems are independent and thus these subproblems can be solved separately. Since constraint $\sum_{k=0}^{K}\sum_{m=1}^{M} d_{m,k}^{1,n} = 1$ and binary variables $d_{m,k}^{1,n} \in \{0,1\}$ in each subproblem in (11), the optimal solution of $D_1^n(\lambda_m, \mu_k)$ is that subcarrier n is allocated exclusively to the node pair (m^*, k^*) such that

$$[m^*, k^*] = \arg\min_{m,k} \; p_m^{1,n} - (\lambda_m - \mu_k) R_{m,k}^{1,n}$$

$$= \arg\min_{m,k} [p_m^{1,n} - (\lambda_m - \mu_k) R_{m,k}^{1,n}] \Big|_{p_m^{1,n} = \max(0, \frac{\lambda_m - \mu_k}{\log 2} - \Gamma W N_0 / |\gamma_{m,k}^n|^2)} \quad (12)$$

and power allocated to subcarrier n is

$$p_{m^*}^{1,n} = \max(0, \frac{\lambda_{m^*} - \mu_{k^*}}{\log 2} - \Gamma W N_0 / \left|\gamma_{m^*,k^*}^n\right|^2) \quad (13)$$

Then we set $d_{m^*,k^*}^{1,n} = 1$ and $d_{m,k}^{1,n} = 0, \forall m \neq m^*, k \neq k^*$. In (12) and (13), $k^* = 0$ means that subcarrier n is allocated to the link between MS m^* and BS in the first uplink frame. To solve (12), $M*(K+1)-1$ comparisons are needed.

Similarly, $D_2(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\varepsilon})$ can be further written as

$$D_2(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\varepsilon}) = \sum_{n=1}^{N} \min_{\substack{\mathbf{P}_m^{2,n}, \mathbf{P}_{k,BS}^{2,n}, \\ \sum_{k=1}^{K} d_k^{2,n} + \sum_{m=1}^{M} d_m^{2,n}=1}} \sum_{m=1}^{M}\sum_{k=0}^{K} d_m^{2,n}[p_m^{2,n} - \lambda_m R_{m,BS}^{2,n}] + d_k^{2,n}[\varepsilon_k p_{k,BS}^{1,n} - \mu_k R_{k,BS}^{1,n}]$$

$$= \sum_{n=1}^{N} D_2^n(\lambda_m, \varepsilon_k, \mu_k)$$

and thus be further decomposed to N subproblems,

$$D_2^n(\lambda_m,\varepsilon_k,\mu_k) = \min_{\substack{\mathbf{P_m^{2,n}},\mathbf{P_{k,BS}^{2,n}},\\ \sum_{k=1}^{K} d_k^{2,n}+\sum_{m=1}^{M} d_m^{2,n}=1}} \sum_{m=1}^{M}\sum_{k=0}^{K} d_m^{2,n}[p_m^{2,n}-\lambda_m R_{m,BS}^{2,n}]+d_k^{2,n}[\varepsilon_k p_{k,BS}^{1,n}-\mu_k R_{k,BS}^{1,n}] \quad \forall 1,...,n,...,N \tag{14}$$

Each subproblem in (14) is to determine the assignment of subcarrier n to one of the links MSs-BS and RSs-BS, and the amount of power of MS or RS spent on the subcarrier. The second uplink frame is equivalent to the uplink transmission in a traditional multiuser OFDMA cellular network with $M+K$ users [23]. Similar to the case of (11), those N subproblems can be solved separately. From constraint $\sum_{k=1}^{K} d_k^{2,n}+\sum_{m=1}^{M} d_m^{2,n}=1$ and binary variables $d_k^{2,n}, d_m^{2,n}\in\{0,1\}$ in each subproblem in (14), it follows that each subcarrier should be allocated to either one of MSs or one of RSs. We first find, for subcarrier n, the optimal MS m^* and the corresponding optimal power allocation respectively, in the following fashion

$$m^* = \arg\min_m [p_m^{2,n}-\lambda_m R_{m,BS}^{2,n}]\Big|_{p_m^{2,n}=\max(0,\,\lambda_m/\log 2-\Gamma W N_0/|\gamma_{m,BS}^n|^2)} \tag{15}$$

$$p_{m^*}^{2,n}=\max(0,\lambda_{m^*}/\log 2-\Gamma W N_0/\left|\gamma_{m^*,BS}^n\right|^2) \tag{16}$$

and then find the optimal RS k^* and the corresponding optimal power allocation, respectively, in the following manner

$$k^* = \arg\min_k [\varepsilon_k p_{k,BS}^{1,n}-\mu_k R_{k,BS}^{1,n}]\Big|_{p_k^{2,n}=\min(\max(0,\frac{\lambda_k}{\varepsilon_k\log 2}-\Gamma W N_0/|\gamma_{k,BS}^n|^2),\,P_k^{\max})} \tag{17}$$

$$p_{k^*}^{2,n}=\min(\max(0,\lambda_{k^*}/(\varepsilon_{k^*}\log 2)-\Gamma W N_0/\left|\gamma_{k^*,BS}^n\right|^2),\ P_{k^*}^{\max}) \tag{18}$$

Then from MS m^* and RS k^* we determine the best one to which the subcarrier is allocated, in the following way

1) If $(p_{m^*}^{2,n}-\lambda_{m^*}R_{m^*,BS}^{2,n})<(\varepsilon_{k^*}p_{k^*,BS}^{1,n}-\mu_{k^*}R_{k^*,BS}^{1,n})$, the subcarrier n is assigned to MS m^* with power allocation according to (16), and $d_{m^*}^{2,n}=1$ and $d_k^{2,n}=0,\forall k, d_m^{2,n}=0,\forall m\neq m^*$ are set.
2) If $(p_{m^*}^{2,n}-\lambda_{m^*}R_{m^*,BS}^{2,n})>(\varepsilon_{k^*}p_{k^*,BS}^{1,n}-\mu_{k^*}R_{k^*,BS}^{1,n})$, the subcarrier n is allocated to RS k^* with power allocation according to (18), and $d_{k^*}^{2,n}=1$ and $d_m^{2,n}=0,\forall m, d_k^{2,n}=0,\forall k\neq k^*$ are set.

When the optimal subcarrier assignment policies and power allocation are determined, the values of $T_{m,BS}^{1}{}^*, T_{m,k}^{1}{}^*, T_{m,BS}^{2}{}^*$ and $T_{k,BS}^{2}{}^*$ can be calculated according to the formulas for link layer rate given in Section Ⅱ.C.

D. *Summary of the Algorithm*

We summarize the complete procedure of the whole algorithm in **Algorithm 1 as** follows.

Algorithm 1: Cross-layer Resource Optimization for Inelastic Flows and Power Efficiency of MSs

1) The BS collects service rate request from all MSs and channel state information through uplink control channels at the beginning of the frame. Then the BS initializes the dual variables $\lambda(0)$ $\mu(0)$ and $\varepsilon(0)$
2) Given $\lambda(t)$, $\mu(t)$ and $\varepsilon(t)$ in iteration t, the BS solves 2N per-subcarrier subproblems in (11) and (14) to obtain the optimal subcarrier assignment and the power allocation in the first and second uplink subframes, respectively.
3) The BS calculates $T_{m,BS}^{1}{}^*, T_{m,k}^{1}{}^*, T_{m,BS}^{2}{}^*$ and $T_{k,BS}^{2}{}^*$ based on the results obtained in step 2), calculates subgradients according to Lemma 1, and then updates the dual variables using (8)
4) Return to step 2) until the algorithm converges.
5) The BS broadcasts the resulting subcarrier assignment policies $\mathbf{d}^*$ and power allocation $\mathbf{P}_{\mathbf{MS}}^*$ and $\mathbf{P}_{\mathbf{RS}}^*$ to all the MSs and RSs through downlink control channels.

IV. CROSS-LAYER TRADEOFF BETWEEN SERVICE RATE OF ELASTIC FLOWS AND POWER EFFICIENCY OF MSS

In this section, we consider elastic flows in the uplink, which can have flexible service rate. A proper rate control scheme is essential for elastic flows to avoid congestion and fairly utilize the available resource. Network Utility Maximization (NUM) has been developed in [14] [24] to formulate joint resource allocation and rate control as an optimization problem. By solving the NUM problem the algorithms have been developed to allocate the resource (bandwidth and power) available in PHY to achieve optimal rate control for each flow. A typical example for a multi-hop CDMA network is showed in [24]. Through jointly optimizing rate control and power allocation under the NUM framework, transport and physical layers are perfectly balanced, i.e., the resource (power) in physical layer is fully utilized while the sum-utility of source rate is maximized.

In cellular networks where MSs are power- and energy-limited device, power efficiency rate control and resource allocation strategies become important issues. Since the service rate of elastic flow is flexible, we can get another balance between power efficiency of MS and service rate of elastic flow. We first get the fundamental tradeoff between rate and energy consumption from Shannon formula for channel capacity. We assume a basic stationary channel with gain γ, bandwidth W, noise power spectral density N_0 and capacity gap Γ. The capacity of the channel with transmission power P is:

$$C = W*\log(1+P|\gamma|^2/\Gamma W N_0)$$

and the energy needed for transmitting one bit is:

$$E_{bit}=P/C=[(W/P)*\log(1+(P/W)|\gamma|^2/\Gamma N_0)]^{-1}$$

It is easy to find that $E_{bit} \to \ln(2)(\Gamma N_0 / |\gamma|^2)$ as $W/P \to \infty$. This indicates that one can reduce energy per bit by reducing the power (which will lower the transmission rate) for given bandwidth. Accordingly, for elastic flows from MSs one can save energy for transmitting the same amount of data by reducing its service rate (through using lower transmission power level).

The above fact motivates us to develop another cross-layer optimization framework and algorithm to balance the service rate of elastic flows and uplink transmission power of MSs. However, to get the best tradeoff in OFDMA cellular networks with fixed RSs, we need to take into account the optimal uplink service rate, full utilization of the power of fixed RSs (which are not energy limited), properly selecting transmission path and assigning OFDMA subcarriers for relaying links and direct links. To this purpose, we will pose a cross-layer tradeoff optimization problem by incorporating the utility of service rate of elastic flows as well as the power consumption of MSs into the objective of the problem, and then solve it using dual decomposition method..

A. Problem for Trade-off Between Service Rate of Elastic Flows and Power Efficiency of MSs

We assume that the utility function associated with elastic flows in MS *m* is $U_m(S_m)$, where S_m is the total service rate of all elastic flows from MS *m* to the BS. $U_m(S_m)$ is usually assumed to be a concave, non-decreasing and continuously differentiable function of S_m. There are many such utility functions available for different optimization objectives. A class of utility functions for fair resource allocation [25] is

$$U^{\alpha}(z) = \begin{cases} (z)^{1-\alpha}/1-\alpha, & \alpha > 0 \text{ and } \alpha \neq 1 \\ \log(z), & \alpha = 1 \end{cases}$$

which is called α-fairness utility. α is a positive constant and represents the level of fairness. For example, maximizing total utility corresponds to the proportional fairness when $\alpha = 1$ and the max-min fairness as $\alpha \to \infty$ [25].

An ideal objective is to achieve simultaneously both maximization of each utility $U_m(S_m)$ and minimization of the power consumption of each MS. The problem in this case can be formulated as a class of vector (multi-criterion) optimization problem as follows [18]

$$\text{Maximize} \quad \Big[\, U_1(S_1), ..., U_{\mathrm{m}}(S_m), ..., U_{\mathrm{M}}(S_M), (-\sum_{t=1}^{2}\sum_{n=1}^{N} p_1^{t,n}), ..., (-\sum_{t=1}^{2}\sum_{n=1}^{N} p_m^{t,n}), ..., (-\sum_{t=1}^{2}\sum_{n=1}^{N} p_M^{t,n}) \Big]$$

subject to (1) (2) (3) (4) (5) and $\mathrm{S} \geq 0$

where S is the vector of service rate S_m, $\forall m$.

We can not obtain 'optimal solution' for this optimization problem because all the objectives are not possible to satisfy at the same time. However, we can find a tradeoff between the service rate of elastic flows and the power consumption of MSs by using the 'scalarization technique' in [18] and introducing $2M$ parameters to get a linear combination of those objectives. The optimization problem in this case becomes

$$\text{Maximize} \quad \sum_{m=1}^{M}[\,\alpha_m U_{\mathrm{m}}(S_m) \text{-} \beta_m \sum_{t=1}^{2}\sum_{n=1}^{N} p_m^{t,n}] \qquad \textbf{(P2)}$$

Subject to (1) (2) (3) (4) (5) and $\mathrm{S} \geq 0$

where α_m and β_m are the weighting parameters associated with MS *m* to determine the tradeoff between the service rate and the power assumption. α_m can be viewed as the reward earned by the utility $U_{\mathrm{m}}(S_m)$, while β_m can be viewed as the price paid to the power consumed by MS *m*. Different tradeoff can be obtained by varying those parameters. Different from **P1**, $S_m, \forall m$ in **P2** are decision variables that need to be optimized for fixed tradeoff parameters α_m and $\beta_m, \forall m$.

B. Solution via Dual Problem

Following the same argument in Section III and the assumption of concavity of utility function, the dual gap of problem **P2** is also zero. Therefore, strong duality holds and we can solve it by dual decomposition and subgradient methods. The corresponding partial Lagrangian of **P2** can be written as follows

$$\begin{aligned}
&L_2(\mathbf{S}, \mathbf{P_{MS}}, \mathbf{P_{RS}}, \mathbf{d}; \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\varepsilon}) \\
&= \sum_{m=1}^{M}[\,\alpha_m U_{\mathrm{m}}(S_m) - \beta_m \sum_{t=1}^{2}\sum_{n=1}^{N} p_m^{t,n}] \\
&\quad - \sum_{m=1}^{M}\lambda_m (S_m - T_{m,BS}^1 - \sum_{k=1}^{K} T_{m,k}^1 - T_{m,BS}^2) \\
&\quad - \sum_{k=1}^{K}\mu_k (\sum_{m=1}^{M} T_{m,k}^1 - T_{k,BS}^2) - \sum_{k=1}^{K}\varepsilon_k (\sum_{n=1}^{N} p_{k,BS}^n - P_k^{\max}) \\
&= \left\{ \sum_{m=1}^{M} \alpha_m \mathrm{U}_{\mathrm{m}}(S_m) - \lambda_m S_m \right\} \\
&\quad + \sum_{n=1}^{N} \left\{ \sum_{m=1}^{M}\sum_{k=0}^{K} d_{m,k}^{1,n} [(\lambda_m - \mu_k) R_{m,k}^{1,n} - \beta_m p_m^{1,n}] \right\} \\
&\quad + \sum_{n=1}^{N} \left\{ \sum_{m=1}^{M}\sum_{k=0}^{K} d_m^{2,n} [\lambda_m R_{m,BS}^{2,n} - \beta_m p_m^{2,n}] + d_k^{2,n} [\mu_k R_{k,BS}^{1,n} - \varepsilon_k p_{k,BS}^{1,n}] \right\}
\end{aligned} \qquad (19)$$

The dual objective function is given by

$$D_t(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\varepsilon}) = \begin{cases} \max\limits_{\mathbf{S}, \mathbf{P_{MS}}, \mathbf{P_{RS}}, \mathbf{d}} L_2(\mathbf{S}, \mathbf{P_{MS}}, \mathbf{P_{RS}}, \mathbf{d}; \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\varepsilon}) \\ s.t. \quad (1)\,(2) \end{cases} \qquad (20)$$

and the corresponding dual problem is

$$\text{Minimize} \quad D_t(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\varepsilon}) \qquad \textbf{(D2)}$$

Subject to $s.t.\ \boldsymbol{\lambda} \geq 0, \boldsymbol{\mu} \geq 0, \boldsymbol{\varepsilon} \geq 0$

With (19), the dual objective function can be decomposed into *M*+2*N* subproblems as follows

1) service rate control

$$D_3^m(\lambda_m) = \max_{S_m} \alpha_m \mathrm{U}_{\mathrm{m}}(S_m) - \lambda_m S_m, \ \forall m = 1, ..., M \qquad (21)$$

2) subcarrier assignment and power allocation in the first uplink subframe

$$D_1^n(\lambda_m,\mu_k)=\max_{\boldsymbol{p_m^{1,n}},\sum_{k=0}^{K}\sum_{m=1}^{M}d_{m,k}^{1,n}=1}\sum_{m=1}^{M}\sum_{k=0}^{K}d_{m,k}^{1,n}[(\lambda_m-\mu_k)R_{m,k}^{1,n}-\beta_m p_m^{1,n}] \quad (22)$$

$$\forall n=1,...,N$$

3) subcarrier assignment and power allocation in the second uplink subframe

$$D_2^n(\lambda_m,\varepsilon_k,\mu_k)$$

$$=\max_{\substack{\boldsymbol{p_m^{2,n}},\boldsymbol{p_{k,BS}^{2,n}},\\ \sum_{k=1}^{K}d_k^{2,n}+\sum_{m=1}^{M}d_m^{2,n}=1}}\sum_{m=1}^{M}\sum_{k=0}^{K}d_m^{2,n}[\lambda_m R_{m,BS}^{2,n}-\beta_m p_m^{2,n}]+d_k^{2,n}[\mu_k R_{k,BS}^{1,n}-\varepsilon_k p_{k,BS}^{1,n}] \quad (23)$$

$$\forall n=1,...,N$$

Given the dual variable λ_m, we can get the optimal service rate for each MS in (21) as follows

$$S_m=\mathrm{U}_m'^{-1}(\lambda_m/\beta_m),\ \forall m=1,...,M \quad (24)$$

where $U_m'^{-1}(\cdot)$ is the inverse function of derivative of utility function $U(\cdot)$.

The solutions to (22) and (23) are similar to the solutions to (10) and (14) (except weighting parameters α_m and β_m).

Similarly, subgradient method can be used to solve dual problem **D2**, which has a convex objective function (20). For the subgradient at given dual variables, we have the following Lemma**.**

Lemma 2: Considering the convex optimization problem **D2**, and assuming that $S_m^*, T_{m,BS}^{1}{}^*, T_{m,k}^{1}{}^*, T_{m,BS}^{2}{}^*, T_{k,BS}^{2}{}^*$ and $p_{k,BS}^{n}{}^*$ are the optimal solution of maximization in the dual objective function (20), which can be obtained by solving (21), (22) and (23), then

$$g_{t,m}(\lambda_m)=T_{m,BS}^{1}{}^*+\sum_{k=1}^{K}T_{m,k}^{1}{}^*+T_{m,BS}^{2}{}^*-S_m^*,$$

$$h_{t,k}(\mu_k)=T_{k,BS}^{2}{}^*-\sum_{k=1}^{M}T_{m,k}^{1}{}^*\quad \text{, and}$$

$$f_{t,k}(\varepsilon_k)=P_k^{\max}-\sum_{n=1}^{N}p_{k,BS}^{n}{}^*\quad ,$$

are the subgradients of $D_t(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\varepsilon})$ at λ_m,μ_k and ε_k, respectively.

The proof of Lemma 2 is similar to that of Lemma 1 which is presented in the Appendix.

Given the optimal solution $S_m^*, T_{m,BS}^{1}{}^*, T_{m,k}^{1}{}^*, T_{m,BS}^{2}{}^*, T_{k,BS}^{2}{}^*$ and $p_{k,BS}^{n}{}^*$ in the current iteration i, the dual variables for **D2** are updated in the following fashion

$$\begin{cases}\lambda_m(i+1)=\max(0,\lambda_m(i)-\sigma(i)g_{t,m}(\lambda_m(i))),\forall m\\ \mu_k(i+1)=\max(0,\mu_k(i)-\phi(i)h_{t,k}(\mu_k(i))),\forall k\\ \varepsilon_k(i+1)=\max(0,\varepsilon_k(i)-\theta(i)f_{t,k}(\varepsilon_k(i))),\forall k\end{cases} \quad (25)$$

where $\sigma(i)$, $\phi(i)$ and $\theta(i)$ are the step-sizes for λ_m,μ_k and ε_k, respectively, in iteration i.

The dual variables converge to the optimum if step-sizes $\sigma(i)$, $\phi(i)$ and $\theta(i)$ are designed appropriately according to **Theorem 1**. Since strong duality holds, the corresponding primal variables ($\mathbf{S}^*,\mathbf{P}_{\mathrm{MS}}^*,\mathbf{P}_{\mathrm{RS}}^*,\mathbf{d}^*$) are globally optimal variables of primal problem **P2** for optimal dual variables ($\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*,\boldsymbol{\varepsilon}^*$).

C. Summary of the Algorithm

We summarize the complete procedure of the whole algorithm in **Algorithm 2** as follows.

Algorithm 2: Resource Optimization for the Cross-layer Tradeoff between Service Rate of Elastic Flows and Power Efficiency of MSs

1) The BS collects the tradeoff parameters α_m and β_m from all MSs and channel state information through uplink control channels at the beginning of the frame. Then the BS initializes the dual variables with $\boldsymbol{\lambda}(0)$, $\boldsymbol{\mu}(0)$ and $\boldsymbol{\varepsilon}(0)$.
2) Given $\boldsymbol{\lambda}(t)$ in iteration *t*, the BS solves *M* service rate control subproblems in (21) using (24) to find S_m^*, $\forall m=1,...,M$.
3) Given $\boldsymbol{\lambda}(t)$, $\boldsymbol{\mu}(t)$ and $\boldsymbol{\varepsilon}(t)$ in iteration *t*, the BS solves 2*N* per-subcarrier subproblems in (22) and (23) to obtain the optimal subcarrier assignment policies and the power allocation in the first and second uplink subframes, respectively.
4) The BS calculates the $T_{m,BS}^{1}{}^*, T_{m,k}^{1}{}^*, T_{m,BS}^{2}{}^*$ and $T_{k,BS}^{2}{}^*$ with the solution obtained in step 2) and 3), calculates subgradients according to Lemma 2, and then updates the dual variables using (25)
4) Return to step 2) until the algorithm converges.
5) The BS broadcasts the resulting service rates S_m^*, subcarrier assignment policies $\boldsymbol{d}^*$ and power allocation $\mathbf{P}_{\mathbf{MS}}^*$ and $\mathbf{P}_{\mathbf{RS}}^*$ to all the MSs and RSs through downlink control channels.

D. Complexity of the Algorithm

In the above subsections, we have solved optimization problem **P2** for the tradeoff between the service rate of elastic flows and the power efficiency of MSs using subgradient method in its dual domain. The problem is decomposed into *M*+2*N* subproblems, i.e., *M* service rate control subproblems in (21), *N* subcarrier assignment and power allocation subproblems in the first uplink subframe in (22) and *N* subcarrier assignment and power allocation subproblems in the second uplink subframe in (23). The complexity of solving per-subcarrier subproblems in the first and second uplink subframes is $O((K+1)M)$ and $O(K+M)$, respectively. The complexity of solving *M* service rate control subproblems is $O(M)$. Thus the complexity of each iteration is $O((KM+K+M)N+M)$, which is linear in *K*, *M* and *N*, respectively. The complexity of subgradient method is polynomial in the number of dual variables (which is 2*K*+*M* for the **D1'**). So the computational complexity of the whole algorithm is linear in the number of the subcarriers *N*, which is significantly lower than employing the exhaustive search solution to the master primal problem **P2** since the number of subcarrier assignment policies ($\mathbf{d}$) increases exponentially with *N*.

The similar results of complexity to **Algorithm 1** in Section III.D can be obtained. The only difference is that the subproblems (21) are not present in **Algorithm 1.** Thus the complexity of computation for each iteration is $O((KM+K+M)N)$.

V. SIMULATION AND RESULTS

To show the performance of an OFDMA cellular network with fixed relays in the cross-layer optimization framework and algorithms proposed in Sections III and VI, a few simulations are conducted and presented here. In the simulations, we consider a wireless OFDMA cellular network with a coverage of a 2-km radius. The MSs are assumed to have low mobility. The distance between the BS and each RS is about 3/5 of the cell radius. The locations of MSs are randomly generated and evenly distributed over the cell. However, from our simulation experience, too large channel gain would significantly reduce the rate of convergence. Thus we impose some additional limits on the locations of MSs as follows

- the distance between any MS and any RS is not less than 300 m;
- the distance between any MS and the BS is not less than 500 m.

Such restriction is reasonable because we care more about the power consumption of MSs far away from any fixed node (RSs or BS).

We model instantaneous channel gain of each subcarrier in a frame as the multiplication of a deterministic path loss with path loss exponent of 4 and a random Rayleigh fading component. We assume that Rayleigh fading component is independent identical distribution among all OFDMA subcarriers. The other parameters and their values used in the simulation are shown in Tab.1. The results in the following simulations are obtained from the average values of 1000 simulation trials.

TABLE I. Simulation Parameters

Simulation Parameter	Value
Total bandwidth	10MHz
Noise power spectral density	-174 dBm/Hz
Number of subcarriers	64
Bandwidth of subcarrier	156.25 kHz
Number of MSs	24
Number of RSs	3, 6
$P_k^{\max}$	36 dBm
Cell Radius	2 km
Γ	1

A. *Inelastic Flow with Fixed Service Rate*

In the first simulation, we assume that there are only inelastic flows with fixed service rate requirement which is the same for all MSs. The average transmission power (mW) consumed per MSs versus the required service rate (Mbps) of the inelastic flow in each MS is showed in Fig. 6 for the scenarios where the numbers of RSs are 3 and 6, respectively. Minimal total power assumption is obtained using **Algorithm1**. The figure shows that the required average power per MS increases with the required service rate. The power consumed by the MS for the 6 RSs case is much less than for the 3 RSs case. This performance gain is about 32%-58% when the number of RSs increases from 3 to 6. This demonstrates that significant reduction of power consumption of MSs in OFMDA cellular network due to the deployment of more RSs can be fully obtained through our cross-layer optimization **Algorithm1.**

Fig. 7 shows the total throughput via RSs versus the total required service rate of inelastic flows from all the MSs (the required service rate for inelastic flows in all MSs is the same). The figure reveals that with the increase of total required service rate, the throughput via RSs increases first for the lower total required service rate until it reaches the maximum and then decreases for the larger total required service rate. The reasons for this are the following

- When the required service rate lies in the lower value range, it is efficient for the MSs that are near to RSs to transmit data by *relaying*. When the required service rate continues to increase, the RSs need to increase their capacity for relaying by utilizing their maximum power and acquiring more subcarriers in the second subframe.
- When the required service rate lies in the higher value range, it is more efficient for some MSs to transmit data to the BS in both subframes (as showed and discussed in the example in Section III.A). Consequently, as the required service rate further increases, more subcarriers in the second subframe are allocated to MSs, which in turn reduces the capacity of relays.

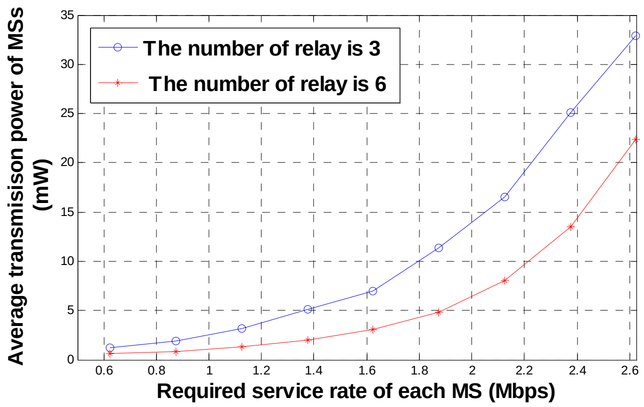

Fig.6 Required service rate versus average power consumption per MS

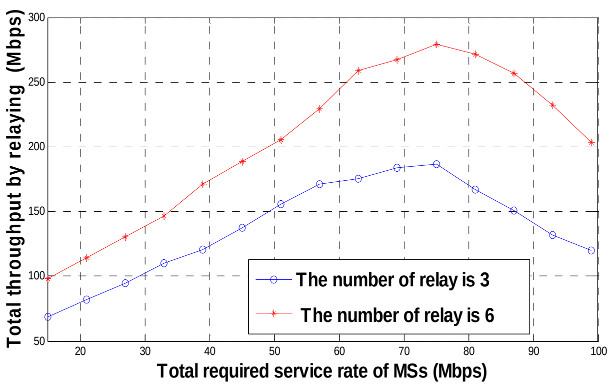

Fig.7 Total required service rate versus total throughput via RSs

B. *Tradeoff between Service Rate of Elastic Flows and Power Efficiency of MSs*

In the second simulation, we show the resource allocation optimization for the tradeoff between the service rate of flows and the power efficiency of MSs.in the case of only elastic flows present. We choose to use the logarithmic utility function $U_{m}(S_m)=\log(S_m), \forall m$, which provides proportional fairness for all elastic flows [25]. To be easy to illustrate the results, we further assume the weighting parameters $\alpha_m = \alpha$ and $\beta_m = \beta, \forall m$ in **P2**. Thus, the ratio β/α can be used to determine the tradeoff between the service rate of flows and the power consumption of MSs. By varying β/α, we examine the change of the average service rate, total utility and average power spent per MS in the cell for two cases where the numbers of RSs are 3 and 6, respectively. The results are obtained based on **Algorienm2** and showed in Figs. 8-10, respectively.

From those figures, we can see that the average service rate, the total utility and the average power consumption decrease with the increase of the value β/α. As mentioned in Section Ⅳ, α_m can be viewed as the reward earned by the utility $U_{m}(S_m)$, while β_m can be viewed as the price paid to the power expense of MS m. Thus increasing β/α implies that the price of power of MSs becomes higher, and this will impose the MSs to reduce power expenditure. This indicates that, for elastic flows we can reduce transmission power of MSs by increasing ratio of the weighting parameter β/α (Fig. 10), which, however, will result in lower average service rate of elastic flows (Fig. 8). But the benefit is that the decrease of the energy required to send one bit data for MSs is decreased (Fig. 11). This means that an MS, which is energy-limited device, can send more data for elastic flows for given fixed total energy in its battery (but with low rate). For each given (β/α), the tradeoff is optimal because of the cross-layer optimization **Algorithm2** in Section III.

Another observation from the results showed in Figs. 8-10 is that more relays deployed can achieve better tradeoff. In other words, the average service rate and total utility in the case of 6 RSs are larger than in the case of 3 RSs, while the average power consumed by MSs is less. But the performance gain in power consumption, which is about 22%-25% (Fig. 10), is smaller than the case of inelastic flows (Fig. 6). This is because we have obtained higher gain in average service rate (Fig. 8) for the 6 RSs case in the tradeoff optimization.

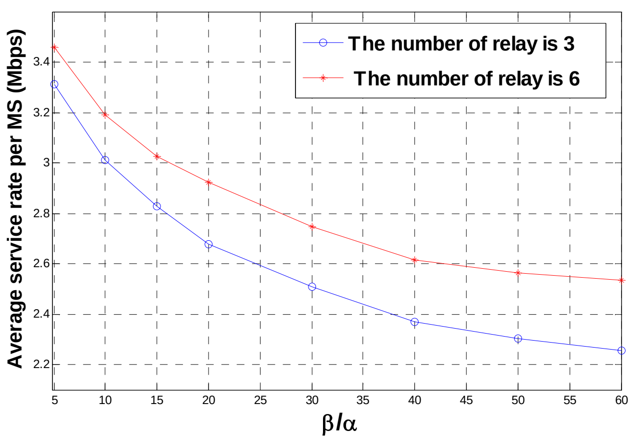

Fig.8 Average service rate versus ratio of the parameters (β/α)

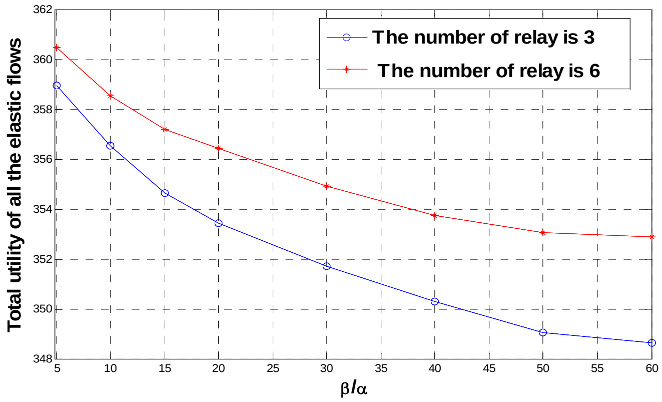

Fig.9 Total utility versus ratio of the parameters (β/α)

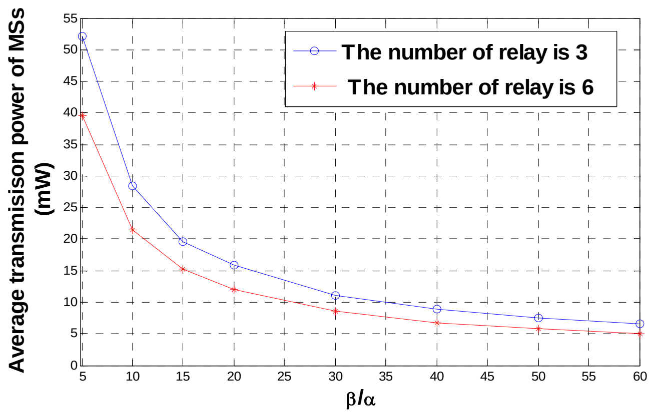

Fig.10 average power consumption per MS versus ratio of the parameters (β/α)

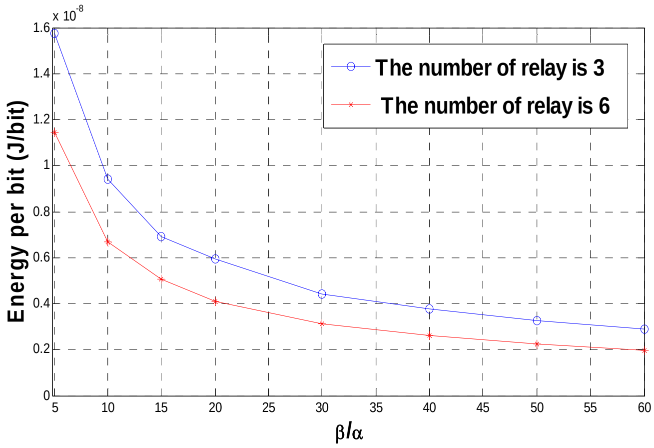

Fig.11 Energy needed for the transmission of one bit versus ratio of the parameters (β/α)

Comparing the results in Figs. 6, 8 and 10 we can notice that it takes much lower average power in the tradeoff optimization of elastic flows to achieve the same average rate as in the optimization of inelastic flow. This is because we impose the strict fairness (i.e., identical service rate) for all the inelastic flows. The MSs being located far away from any RS or BS have to spend quite a lot of power to satisfy this rate requirement, which results high average power consumption. While in the tradeoff optimization of its elastic flow, such MSs may lower their service rate to reduce its transmission power, and thus the average power consumption is reduced.

VI. Conclusion

In this paper, we have developed the cross-layer resource optimization framework for both inelastic flows and elastic flows in the uplink transmission of an OFDMA cellular network with fixed RSs. The RSs are assumed to be dedicated devices to relay the data from MSs to a BS and have no energy limitation while the MSs are power- and energy-limited devices. We have formulated the cross-layer optimization problem to minimize the sum power consumption of MSs in the case of inelastic flow and presented the cross-layer tradeoff between the service rate and power consumption of MSs in the case of elastic flows. Dual decomposition and subgradient update method ware employed to obtain optimal solution with reduced computational complexity. Simulation results showed that through the proposed cross-layer resource optimization framework and algorithms, the benefit of deployment of multiple RSs in the uplink transmission of OFDMA cellular network in significantly reducing power consumption in inelastic flow case can be fully obtained and tradeoff between benefits in increasing service rate and saving energy can also be achieved.

VII. Acknowledgements

In this work Lei You is financially supported by VINNOVA Sweden through grant 2008-00954 and Chinese Schoolarship Council (CSC), and Ping Wu is supported by VINNOVA. Mei Song, Junde Song and Yong Zhang are financed by the Ministry of Science and Technology (MOST) China through National International Science and Technology Cooperation Project (NO. 2008DFA12090).

References

[1] Pabst R,Walke H B,Schultz C D,et al. "Relay-Based Deployment Concepts for Wireless and Mobille Broadband Radio", IEEE Communication Magazine,vol.42, No.9, pp.80–89, Sep.2004.

[2] IEEE 802.16's Relay Task Group, http://ieee802.org/16/relay/index.htm

[3] Rhee, W., Cioffi, J.M. "Increasing in capacity of multiuser OFDM system using dynamic subchannel allocation,". In Proceedings of Vehicular Technology Conference, VTC2000-Spring, 15-18 May, 2000, pp.1085–1089, vol.2

[4] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE J. Sel. Areas Commun.,* vol. 21, pp.171-178, Feb. 2003.

[5] Y.W. Cheong, R.S. Cheng, K.B. Lataief, and R.D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," IEEE J. Select. Areas Commun., vol. 17, no. 10, pp. 1747-1758, Oct. 1999.

[6] Lei Huang, Mengtian Rong, Lan Wang, et al. "Resource Allocation for OFDMA Based Relay Enhanced Cellular Networks", In Proceedings of Vehicular Technology Conference, VTC2007, pp. 3160–3164, April 2007.

[7] Guoqing Li, Hui Liu. "Resource Allocation for OFDMA Relay Networks With Fairness Constraints," IEEE Journal on Selected Areas in Communications, Volume 24, Issue 11, pp.2061 – 2069, Nov. 2006.

[8] R. Cendrillon, W. Yu, M. Moonen, J. Verlinden, and T. Bostoen, "Optimal multiuser spectrum balancing for digital subscriber lines," IEEE Trans. Commun., vol. 54, no. 5, pp. 922–933, May 2006.

[9] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," IEEE Trans. Commun., vol. 54, no. 7, pp. 1310– 1322, Jul. 2006.

[10] Kibeom Seong, Mohseni, M., Cioffi, J.M.. "Optimal Resource Allocation for OFDMA Downlink Systems**.**" in Proceedings of IEEE International Symposium on Information Theory, 2006. 9-14 July 2006 Page(s):1394 – 1398.

[11] T. Ng, and W. Yu, "Joint Optimization of Relay Strategies and Resource Allocations in Cooperative Cellular Networks," IEEE Journal on Selected Areas in Communications, Vol.25, No. 2, pp. 328-339 Feb.2007

[12] Nam, W, Chang, W, Chung, S.-Y, Lee, Y. H. "Transmit Optimization for Relay-Based Cellular OFDMA Systems", in Proceedings of IEEE ICC'07, June 2007, pp. 5714-5719.

[13] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," J. Operations Research Society, vol. 49, no. 3, pp. 237–252, 1998.

[14] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures," Proc. IEEE, vol. 95. no. 1, pp. 255-312, Jan. 2007.

[15] Daniel P. Palomar and Mung Chiang, "A Tutorial on Decomposition Methods for Network Utility Maximization," IEEE Journal on Selected Areas in Communications, Vol. 24, No. 8, pp. 1439-1451, Aug. 2006.

[16] Ki-Dong Lee, Leung, V.C.M, "Evaluations of Achievable Rate and Power Consumption in Cooperative Cellular Networks With Two Classes of Nodes", IEEE Transactions on Vehicular Technology. Volume 57, Issue 2, pp. 1166 – 1175, March 2008.

[17] L.M.C. Hoo, B. Halder, J. Tellado, and J.M. Cioffi, "Multiuser Transmit Optimization for Multicarrier Broadcast Channels: asymptotic FDMA capacity region and algorithms," IEEE Trans. Comm., vol. 52, no. 6, pp. 922-930, June 2004.

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

[19] Kemal Karakayali, Kang, J.H., Murali Kodialam, et al. "Cross-Layer Optimization for OFDMA-Based Wireless Mesh Backhaul Networks", in Proceedings of IEEE Wireless Communications and Networking Conference, 2007. 11-15 March 2007, pp. 276 – 281.

[20] Soldati, P.; Johansson, M., "Network-Wide Resource Optimization of Wireless OFDMA Mesh Networks with Multiple Radios". in Proceedings of IEEE International Conference on Communications, 2007. ICC '07. 24-28 June 2007, pp. 4979 – 4984.

[21] D. P. Bertsekas, Nonlinear Programming, 2nd ed. Belmont, MA: Athena Scientific, 1999.

[22] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," in lecturenotes of EE392o, Stanford University, Autumn Quarter 2003-2004.

[23] Keunyoung Kim, Youngnam Han, Seong-Lyun Kim. "Joint subcarrier and power allocation in uplink OFDMA systems". IEEE Communications Letters, Volume 9, Issue 6, pp. 526–528, Jun 2005.

[24] M. Chiang, "Balancing transport and physical layers in wireless multihop networks: Jointly optimal congestion control and power control," IEEE J. Sel. Areas Commun., vol. 23, no. 1, pp. 104–116, Jan. 2005.

[25] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," IEEE/ACM Trans. Networking, vol. 8, no. 5, pp. 556–567, Oct. 2000.

[26] J. Huang, V. Subramanian, R. Berry and R. Agrawal, "Joint Scheduling and Resource Allocation in Uplink OFDM Systems," Proc. of Asilomar Conference, Pacific Grove, CA, Nov. 2007.

Appendix

Proof of lemma 1

Proof: Definition of subgradient [21] [22]-given a convex function $f: R^n \to R$, a vector $h \in R^n$ is a subgradient of f at the point $v \in R^n$ if $f(u) \ge f(v) + (u-v)^T h$, $\forall u \in R^n$.

Consider objective function $-D(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\varepsilon})$ in **D1'** at two different points $(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\varepsilon})$ and $(\boldsymbol{\lambda}',\boldsymbol{\mu},\boldsymbol{\varepsilon})$, where $\boldsymbol{\lambda} = (\lambda_1,...,\lambda_m,...,\lambda_M)$ and $\boldsymbol{\lambda}' = (\lambda_1,...,\lambda_m',...,\lambda_M)$. We have

$$-D(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\varepsilon}) = \begin{cases} -\min\limits_{\mathbf{P_{MS}},\mathbf{P_{RS}},\mathbf{d}} L_1(\mathbf{P_{MS}},\mathbf{P_{RS}},\mathbf{d};\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\varepsilon}) \\ s.t. \quad (1)\,(2) \end{cases} \tag{26}$$

$$-D(\boldsymbol{\lambda}',\boldsymbol{\mu},\boldsymbol{\varepsilon}) = \begin{cases} -\min\limits_{\mathbf{P_{MS}},\mathbf{P_{RS}},\mathbf{d}} L_1(\mathbf{P_{MS}},\mathbf{P_{RS}},\mathbf{d};\boldsymbol{\lambda}',\boldsymbol{\mu},\boldsymbol{\varepsilon}) \\ s.t. \quad (1)\,(2) \end{cases} \tag{27}$$

Letting the optimal values of $\mathbf{P_{MS}},\mathbf{P_{RS}}$ and $\mathbf{d}$ in (26) and (27) be $\mathbf{P}^*_{\mathbf{MS}},\mathbf{P}^*_{\mathbf{RS}},\mathbf{d}^*$ and $\mathbf{P}'^*_{\mathbf{MS}},\mathbf{P}'^*_{\mathbf{RS}},\mathbf{d}'^*$, respectively, we can find the subgradient of $-D(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\varepsilon})$ at λ_m in the following manner

$$\begin{aligned}
&[-D(\boldsymbol{\lambda}',\boldsymbol{\mu},\boldsymbol{\varepsilon})] - [-D(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\varepsilon})] \\
&= -L_1(\mathbf{P}'^*_{\mathbf{MS}},\mathbf{P}'^*_{\mathbf{RS}},\mathbf{d}'^*;\boldsymbol{\lambda}',\boldsymbol{\mu},\boldsymbol{\varepsilon}) + L_1(\mathbf{P}^*_{\mathbf{MS}},\mathbf{P}^*_{\mathbf{RS}},\mathbf{d}^*;\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\varepsilon}) \\
&\ge -L_1(\mathbf{P}^*_{\mathbf{MS}},\mathbf{P}^*_{\mathbf{RS}},\mathbf{d}^*;\boldsymbol{\lambda}',\boldsymbol{\mu},\boldsymbol{\varepsilon}) + L_1(\mathbf{P}^*_{\mathbf{MS}},\mathbf{P}^*_{\mathbf{RS}},\mathbf{d}^*;\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\varepsilon}) \\
&= -\lambda_m'(S_m - T^1_{m,BS}{}^* - \sum_{k=1}^{K} T^1_{m,k}{}^* - T^2_{m,BS}{}^*) + \lambda_m(S_m - T^1_{m,BS}{}^* - \sum_{k=1}^{K} T^1_{m,k}{}^* - T^2_{m,BS}{}^*) \\
&= (\lambda_m' - \lambda_m)(T^1_{m,BS}{}^* + \sum_{k=1}^{K} T^1_{m,k}{}^* + T^2_{m,BS}{}^* - S_m)
\end{aligned} \tag{28}$$

$T^1_{m,BS}{}^*$, $T^1_{m,k}{}^*$ and $T^2_{m,BS}{}^*$ are the link layer rate of links MS m-BS and MS m-RS k in the first uplink subframe and MS m-BS in the second uplink subframe, respectively，when the dual variables are $(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\varepsilon})$. The inequality in (28) holds because the definition of dual function in (7). The second equality holds because the definition of Lagrange in (6). Thus, we get

$$-D(\boldsymbol{\lambda}',\boldsymbol{\mu},\boldsymbol{\varepsilon}) \ge -D(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\varepsilon}) + (\lambda_m' - \lambda_m)(T^1_{m,BS}{}^* + \sum_{k=1}^{K} T^1_{m,k}{}^* + T^2_{m,BS}{}^* - S_m)$$

By the definition of subgradient, the subgradient of $-D(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\varepsilon})$ at the point λ_m is

$$g_m(\lambda_m) = T^1_{m,BS}{}^* + \sum_{k=1}^{K} T^1_{m,k}{}^* + T^2_{m,BS}{}^* - S_m$$

Similarly, we can also get the subgradients of $-D(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\varepsilon})$ at μ_k and ε_k, which are , respectively, of the following form

$$h_k(\mu_k) = T^2_{k,BS}{}^* - \sum_{k=1}^{M} T^1_{m,k}{}^* \quad ,$$

$$f_k(\varepsilon_k) = P_k^{\max} - \sum_{n=1}^{N} p^n_{k,BS}{}^* \quad .$$

□