

Research Article

Fair Adaptive Bandwidth and Subchannel Allocation in the WiMAX Uplink

Antoni Morell, Gonzalo Seco-Granados, and José López Vicario

Telecommunications and System Engineering Department (TES), Autonomous University of Barcelona (UAB), 08193 Bellaterra, Spain

Correspondence should be addressed to Antoni Morell, antoni.morell@uab.cat

Received 2 July 2008; Revised 22 November 2008; Accepted 29 December 2008

Recommended by Ekram Hossain

In some modern communication systems, as it is the case of WiMAX, it has been decided to implement Demand Assignment Multiple Access (DAMA) solutions. End-users request transmission opportunities before accessing the system, which provides an efficient way to share system resources. In this paper, we briefly review the PHY and MAC layers of an OFDMA-based WiMAX system, and we propose to use a Network Utility Maximization (NUM) framework to formulate the DAMA strategy foreseen in the uplink of IEEE 802.16. Utility functions are chosen to achieve fair solutions attaining different degrees of fairness and to further support the QoS requirements of the services in the system. Moreover, since the standard allocates resources in a terminal basis but each terminal may support several services, we develop a new decomposition technique, the coupled-decompositions method, that obtains the optimal service flow allocation with a small number of iterations (the improvement is significant when compared to other known solutions). Furthermore, since the PHY layer in mobile WiMAX has the means to adapt the transport capacities of the links between the Base Station (BS) and the Subscriber Stations (SSs), the proposed PHY-MAC cross-layer design uses this extra degree of freedom in order to enhance the network utility.

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

1. Introduction

The wireless community has recently directed much attention on a variety of topics related to Worldwide Interoperability for Microwave Access (WiMAX) technologies as a broadband solution. Two different standards are under this commercial nomenclature: the IEEE 802.16 [1], with its extension to mobile scenarios IEEE 802.16e [2], and the ETSI HiperMAN [3]. Operating in the range of 2 GHz to 11 GHz, WiMAX enables a fast deployment of the network even in remote locations with low coverage of wired technologies, such as the Digital Subscriber Loop (DSL) family, and it can be used, among others, for wireless backhaul or last-mile applications.

The IEEE 802.16 standards family provides manufacturers with basically four different physical (PHY) layers [4]. Two of them are based on single carrier transmissions and use Time Division Multiple Access (TDMA) whereas the other two are based on multicarrier modulations and use either TDMA or Orthogonal Frequency Division Multiple Access (OFDMA). Within the multicarrier subgroup, the

WirelessMAN Orthogonal Frequency Division Multiplexing (OFDM) uses a 256-point Fast Fourier Transform- (FFT-) based OFDM modulation together with a TDMA scheme to deploy a Point-to-Multipoint (PMP) subnetwork in the frequency range from 2 GHz up to 11 GHz in Non-Line-of-Sight (NLOS) propagation conditions. This PHY layer has been accepted for fixed WiMAX applications, and it is often termed as fixed WiMAX. Finally, WirelessMAN OFDMA exploits the multicarrier principles to implement a more flexible OFDMA access scheme. As in WirelessMAN OFDM, it is intended for NLOS PMP applications in the 2 GHz–11 GHz range. However, it uses a variable-size FFT ranging from 128 up to 2048 subcarriers. This PHY layer has been accepted for mobile WiMAX applications, and it is usually termed mobile WiMAX.

Concerning network topology, the basic configuration is PMP with a Base Station (BS) serving many Subscriber Stations (SSs). Not with standing, there is also a mesh mode available where SSs can be linked directly to the BS or routed through other SSs. This last mode is out of the scope of this paper, where we consider the design

of appropriate scheduling mechanisms in uplink using the WirelessMAN OFDMA PHY layer and a PMP network. The conceived scheduling mechanism is based on a Demand Assignment Multiple Access (DAMA) strategy that implements a Dynamic Bandwidth Allocation (DBA) solution (where bandwidth is understood as rate in a wide sense). Jointly with flow allocation, we consider the adjustment of the transmission parameters of the OFDM system, and hence, the joint approach proposes a cross-layer interaction between PHY and Medium Access Control (MAC) system layers.

Previous works related to Radio Resource Management (RRM) in WiMAX networks address a variety of scenarios, from PMP to mesh, from TDMA to OFDMA access types, and distinguishing single channel or multichannel networks, most of them from a physical (PHY) layer perspective, where the goal is to properly configure the transmission parameters. At the best of our knowledge, two main approaches are found in literature, namely: (i) formulate the problem in a mathematical optimization framework and (ii) develop heuristic algorithms. In the sequel, we briefly review some of the works. In [5], the author proposes an heuristic solution for the case of a single cell OFDMA WiMAX network that maximizes the network sum-rate under some fairness considerations by means of performing subcarrier and power allocation. The authors in [6] analyze how concurrent transmissions boost performance in mesh type networks by proposing an interference-aware routing and scheduling mechanism. In [7], the reader can find a discussion about the advantages of a multichannel network. Finally, [8] contributes with a mathematical optimization solution that falls into the Network Utility Maximization (NUM) framework, where a distributed optimal solution to the established NUM problem is obtained using a convex decomposition approach. The authors extend in [9] their original work to generic OFDMA mesh networks, and the contributions in [10–12] are within the same context. A common feature in the last three references is that they split the global rate control and resource allocation problem into independent and smaller subproblems in order to alleviate the complexity of the solution at the expenses of a certain loss in optimality.

Our work follows the NUM framework to define the underlying optimization problem as in [8] but modifies the formulation in order to exactly fit the DAMA process that is envisaged for the WiMAX uplink. The problem is then decomposed (without any loss in optimality) using the Mean Value Cross (MVC) decomposition method [13]. It allows to separate the original joint problem into a flow optimization problem (given fixed link capacities) and a radio resource optimization problem (given fixed values of transmission rates). The latter results in a linear program that can be solved centrally at the BS, whereas a distributed solution that uses the novel proposed coupled-decompositions method is applied to the former.

The rest of the paper is organized as follows. Section 2 describes the system model. Section 3 reviews the MVC decomposition technique and introduces the novel coupled-decompositions method, whereas Section 4 solves the pro-

posed joint problem in Section 2. Finally, Section 5 gives some numerical results, and Section 6 ends the paper with the conclusions.

2. System Model

Let us consider a PMP OFDMA WiMAX network as depicted in Figure 1, where a number of SSs share a subset of the subchannels in the system. A subchannel in WiMAX is made up of some of the system subcarriers and lasts for several OFDM symbols in time. There exist different ways to gather subcarriers into subchannels, which depend on the permutation types (see in [4] a good review on WiMAX aspects). In this work, we assume that the transmitting power per subchannel as well as the set of subcarriers that form it is given. Therefore, the different powers are not variables of our allocation problem. Furthermore, each terminal allocates the amount of power at each subchannel among the inner subcarriers in order to optimize the transmitting rate. This assumption can be found in [14], where the authors take into account intercell interference to constrain the subchannel transmitting powers. Note that one interesting extension is then the inclusion of subchannel power allocation but it is beyond the scope of this paper. In our framework, given a specific allocation of subchannels to terminals $\{\rho_i\}$ (top left part of the figure), each terminal is able to transmit at a rate $c^i(\rho_i)$, which is the sum of the rates that the SS attains in its active subchannel subset (the subset allocated to the terminal).

We further assume (as described in the IEEE 802.16 standard documents) that each terminal negotiates the resource allocation for all traffic flows that go through it, that is, it jointly requests transmission opportunities for the ongoing connections without doing it on a flow basis. The advantage of this procedure is that signaling is reduced, specially when a significant number of connections have to be managed. The disadvantage is that, depending on the particular mechanism used to find the solution of the problem, it may not be optimal. In that sense, solutions derived from distributed optimization do not sacrifice optimality. The price to pay is the time required to get the solution, and therefore, we are interested in techniques that converge fast. In Figure 1, the rate of the j th flow at the i th SS is labeled as r_j^i .

The IEEE 802.16 standard defines five different scheduling services that will provide Quality of Service (QoS) differentiation among the multiple traffic types. These services are [4] (i) the Unsolicited Grant Service (UGS) (ii) the real-time Polling Service (rtPS) (iii) the non-real-time Polling Service (nrtPS) (iv) the Best-Effort (BE) service, and (v) the extended real-time Polling Service (ertPS). Let us model the DAMA solution implemented in the WiMAX uplink by means of a convex program [15] where the different scheduling services are mapped using three parameters: the minimum rate that has to be allocated to the connection (the j th flow at the i th terminal) or m_j^i , the rate requested or d_j^i , and the priority of the service or p_j^i . The desired QoS degree of each service depends then on both m_j^i and p_j^i . For example, the UGS that needs a constant rate can be requested just by plugging

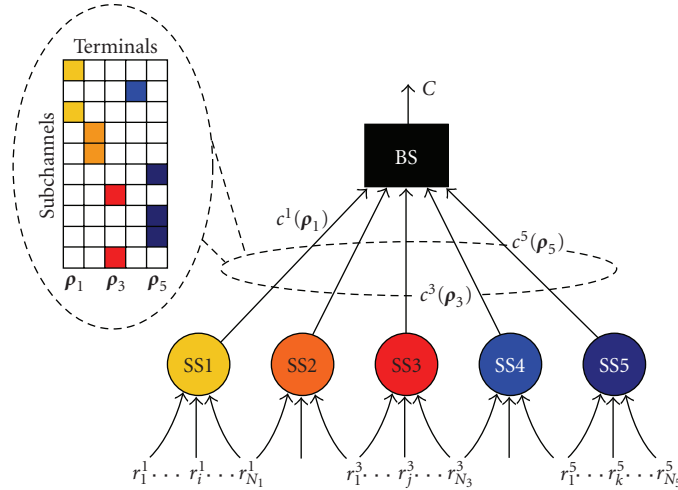


FIGURE 1: Reference model.

that rate into m_j^i and fixing $d_j^i = m_j^i$ regardless the value of p_j^i . Another example is the ertPS that can be requested with some amount of m_j^i for the fixed allocation part and some $d_j^i > m_j^i$ for the variable rate part. The value p_j^i is then used to prioritize this flow against other competing connections.

In summary, the cross-layer system model used to characterize the DBA part of WiMAX, including PHY and MAC layer issues, responds to the following convex optimization problem in maximization form [15, Section 4.1.3]:

$$\begin{aligned}
 & \max_{\{r_j^i\}, \Gamma} \sum_{i=1}^N \sum_{j=1}^{N_i} U_j^i(r_j^i; p_j^i) \\
 & \text{s.t.} \quad \sum_{i=1}^N \sum_{j=1}^{N_i} r_j^i \leq C, \\
 & \sum_{j=1}^{N_i} r_j^i \leq c^i(\rho_i), \quad i = 1, \dots, N, \\
 & m_j^i \leq r_j^i \leq d_j^i, \quad \forall i, \forall j, \\
 & \Gamma \mathbf{1} \preceq \mathbf{1}, \\
 & \rho_i \succeq \mathbf{0}, \quad i = 1, \dots, N,
 \end{aligned} \tag{1}$$

where $U_j^i(r_j^i; p_j^i)$ is the function that measures the utility perceived by the connection when the rate r_j^i is allocated. The function has p_j^i as a parameter. Furthermore, $\Gamma = [\rho_1, \dots, \rho_N]$ collects the subchannel allocation per user (ρ_i), and the symbols \preceq and \succeq stand for component-wise non-strict inequalities. Finally, $c^i(\rho_i) = \rho_i^T \mathbf{c}_i$, where \mathbf{c}_i contains the achievable rates of SS $_i$ at each possible subchannel, and C is the rate at which the BS can transmit. Note that in principle the allocation variables within each vector ρ_i should take the integer values 0 and 1 so that a given subchannel is completely allocated to a certain SS, whereas the constraint $\Gamma \mathbf{1} \preceq \mathbf{1}$ forces that no more than one terminal gets the subchannel. As it has been done in other works in literature [16], we relax the integer constraints to $\rho_i^k \geq 0$, which allows us to represent the problem as a convex one (easy to

solve). Once the solution of the relaxed problem is found, a suboptimal solution to the original problem (with integer constraints) is obtained by means of employing rounding algorithms. However, in the WiMAX scenario and taking into account that an allocation is kept during several time-slots, real-valued allocation variables have sense in practice (by time sharing of subchannels). Indeed, if we consider that the allocation lasts for T time slots, then it is possible to use values in Γ with a granularity of $1/T$.

Not with standing, the problem in (1) itself does not guarantee a fair allocation of resources. Fortunately, such distribution can be attained by means of employing adequate utility functions, and a general formulation for fairness was introduced in [17] under the nomenclature of (\mathbf{p}, α) -proportional fairness. A feasible rate vector \mathbf{r}^\dagger (i.e., it attains the generic network constraints $\mathbf{A}\mathbf{r}^\dagger \preceq \mathbf{c}$) is said to be (\mathbf{p}, α) -proportionally fair (where $\mathbf{p} = [p_1, \dots, p_{N'}]^T$ and α are positive real numbers) if, given any other feasible rate vector \mathbf{r}^\ddagger , it holds that

$$\sum_{i=1}^{N'} p_i \frac{r_i^\ddagger - r_i^\dagger}{(r_i^\dagger)^\alpha} \leq 0, \quad \forall \mathbf{r}^\ddagger \text{ s.t. } r_i^\ddagger \geq 0, \mathbf{A}\mathbf{r}^\ddagger \preceq \mathbf{c}. \tag{2}$$

Accordingly, the utility functions that accomplish this fairness criterion are [17]

$$U_i(r_i; p_i, \alpha) = \begin{cases} p_i \log(r_i), & \alpha = 1, \\ p_i \frac{r_i^{(1-\alpha)}}{1-\alpha}, & \alpha \neq 1. \end{cases} \tag{3}$$

The reader can find in Figure 2 the plots of $U_i(r_i; p_i, \alpha)$ for $\alpha = 0.1$, $\alpha = 1$, and $\alpha = 3$ (equal p_i value).

Let us fix $\mathbf{p} = [1, \dots, 1]^T$ and move from $\alpha \rightarrow \infty$ to $\alpha = 0$. With $\alpha \rightarrow \infty$, the solution is said to be max-min fair [18, Section 6.5], and it is not possible (given feasibility, i.e., $\mathbf{A}\mathbf{r} \preceq \mathbf{c}$) to increase any rate in the network, say r_j , without decreasing another rate $r_p < r_j$. On the other hand, when $\alpha \rightarrow 0$, the flow allocation problem leads to a max sum-rate approach, and therefore, it drastically favors the users

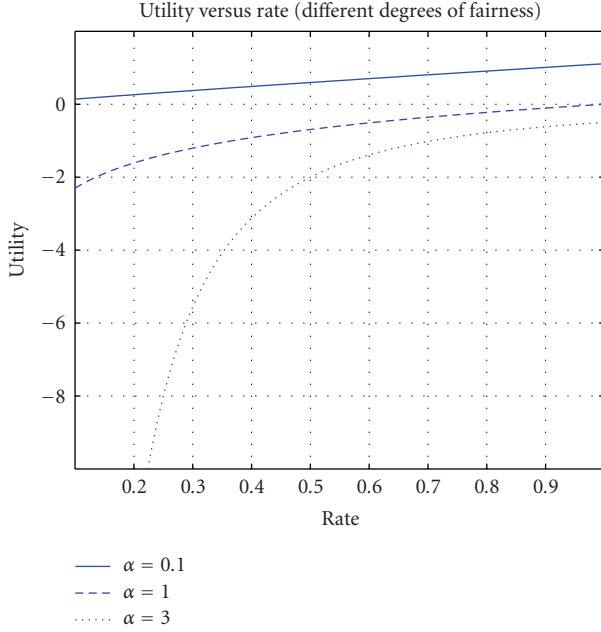


FIGURE 2: Different degrees of fairness (α) in the definition of utility functions.

with better links (it is then unfair). Intermediate solutions allow a certain decrease in r_p at the expenses of a greater increase in r_j depending on α . Note that in Figure 2 the bigger the α value is, the higher the increase in r_j will be in order to compensate a utility loss in r_p . A common adopted solution in literature is $\alpha = 1$, and it was termed by Kelly [19] as proportional fair. Moreover, this solution coincides with the Nash Bargaining one, and therefore, it accomplishes the recognized, axioms in game theory [20] of linearity, irrelevant alternatives and symmetry [21].

We can conclude that there is no unique criterion to define fairness but a series of them are explicitly characterized with the utility functions in (3). Furthermore, some flows can be prioritized over the others within a specific fairness framework (fixed by α) by particular adjustment of the scale thanks to the parameters $\{p_i\}$. In general, proportional fairness ($\alpha = 1$) provides a reasonable trade-off between fairness and resource utilization (network throughput).

3. Decomposition in Convex Programming

Decomposition techniques are used to break down a given optimization problem into a number of smaller problems, usually termed the subproblems. The most used decomposition methods in communications literature and in relation to convex optimization are primal and dual decompositions [22, 23]. It is usual to employ these decomposition techniques as a tool to obtain distributed solutions to some problems, as it is the case in Network Utility Maximization (NUM) problems [24, 25]. The formulation in (1) is an adaptation of the classical NUM to match the DBA problem in OFDMA WiMAX. Recently, Palomar and Chiang provided an exhaustive review on primal and dual decompositions

applied to the classical NUM and extensions of it [26]. In particular, they proposed multilevel decomposition approaches to split the problem into different and coupled subsets of variables (e.g., link powers and transmission rates). However, the problem in primal and dual decompositions is that, in general, they converge slowly and that an adaptation step size has to be fixed by the user. So motivated, we base our work in two distinct decomposition techniques: the Mean Value Cross (MVC) decomposition [13] and the proposed novel coupled-decompositions method. In the following, we briefly review the former and describe the latter.

3.1. Mean Value Cross Decomposition. Consider the following problem formulation from [13]:

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{y}} \quad & c(\mathbf{x}) + e(\mathbf{y}) \\ \text{s.t.} \quad & \mathbf{A}_1(\mathbf{x}) + \mathbf{B}_1(\mathbf{y}) \leq \mathbf{b}_1, \\ & \mathbf{A}_2(\mathbf{x}) + \mathbf{B}_2(\mathbf{y}) \leq \mathbf{b}_2, \\ & \mathbf{x} \in \mathcal{X}, \\ & \mathbf{y} \in \mathcal{Y}, \end{aligned} \quad (4)$$

where $c : \mathbb{R}^{n_1} \rightarrow \mathbb{R}$, $e : \mathbb{R}^{n_2} \rightarrow \mathbb{R}$, $\mathbf{A}_1 : \mathbb{R}^{n_1} \rightarrow \mathbb{R}^{m_1}$, $\mathbf{B}_1 : \mathbb{R}^{n_2} \rightarrow \mathbb{R}^{m_1}$, $\mathbf{A}_2 : \mathbb{R}^{n_1} \rightarrow \mathbb{R}^{m_2}$, and $\mathbf{B}_2 : \mathbb{R}^{n_2} \rightarrow \mathbb{R}^{m_2}$ are convex functions. The sets \mathcal{X} and \mathcal{Y} are also convex and compact. It is further assumed that strong duality holds.

Construct now the partial Lagrangian function of the problem (4) as

$$L(\mathbf{x}, \mathbf{y}, \boldsymbol{\mu}) = c(\mathbf{x}) + e(\mathbf{y}) + \boldsymbol{\mu}^T (\mathbf{A}_1(\mathbf{x}) + \mathbf{B}_1(\mathbf{y}) - \mathbf{b}_1) \quad (5)$$

and minimize it over the variable \mathbf{x} , including the constraints that have not been taken into account in the Lagrangian definition, to obtain the function $K(\mathbf{y}, \boldsymbol{\mu})$ as follows:

$$\begin{aligned} K(\mathbf{y}, \boldsymbol{\mu}) = \min_{\mathbf{x}} \quad & L(\mathbf{x}, \mathbf{y}, \boldsymbol{\mu}) \\ \text{s.t.} \quad & \mathbf{A}_2(\mathbf{x}) \leq \mathbf{b}_2 - \mathbf{B}_2(\mathbf{y}), \\ & \mathbf{x} \in \mathcal{X}, \end{aligned} \quad (6)$$

which is convex in \mathbf{y} and concave in $\boldsymbol{\mu}$ [13].

From $K(\mathbf{y}, \boldsymbol{\mu})$, the method defines the primal and the dual subproblem by fixing either the primal variable \mathbf{y} or the dual variable $\boldsymbol{\mu}$. After some manipulations, the primal subproblem turns into

$$\begin{aligned} p(\mathbf{y}) = \min_{\mathbf{x}} \quad & c(\mathbf{x}) + e(\mathbf{y}) \\ \text{s.t.} \quad & \mathbf{A}_1(\mathbf{x}) \leq \mathbf{b}_1 - \mathbf{B}_1(\mathbf{y}), \\ & \mathbf{A}_2(\mathbf{x}) \leq \mathbf{b}_2 - \mathbf{B}_2(\mathbf{y}), \\ & \mathbf{x} \in \mathcal{X} \end{aligned} \quad (7)$$

and the dual subproblem into

$$\begin{aligned} d(\boldsymbol{\mu}) = \min_{\mathbf{x}, \mathbf{y}} \quad & c(\mathbf{x}) + e(\mathbf{y}) + \boldsymbol{\mu}^T (\mathbf{A}_1(\mathbf{x}) + \mathbf{B}_1(\mathbf{y}) - \mathbf{b}_1) \\ \text{s.t.} \quad & \mathbf{A}_2(\mathbf{x}) + \mathbf{B}_2(\mathbf{y}) \leq \mathbf{b}_2, \\ & \mathbf{x} \in \mathcal{X}, \\ & \mathbf{y} \in \mathcal{Y}. \end{aligned} \quad (8)$$

Finally, the method is completed by passing filtered versions of the primal and dual variables between the primal and dual subproblems, as it is summarized in the following algorithm.

Take starting points $\boldsymbol{\mu}^0 \succcurlyeq \mathbf{0}$ and $\mathbf{y}^0 \in \mathcal{Y}$ and let $k = 1$.

Repeat

- (1) Let $\bar{\boldsymbol{\mu}}^k = (1/k) \sum_{i=0}^{k-1} \boldsymbol{\mu}^{i-1} = (1/k) \boldsymbol{\mu}^{k-1} + ((k-1)/k) \bar{\boldsymbol{\mu}}^{k-1}$ and compute $d(\bar{\boldsymbol{\mu}}^k)$ as in (8). Get \mathbf{y}^k as the inner minimizer of $d(\bar{\boldsymbol{\mu}}^k)$.
- (2) Let $\bar{\mathbf{y}}^k = (1/k) \sum_{i=0}^{k-1} \mathbf{y}^{i-1} = (1/k) \mathbf{y}^{k-1} + ((k-1)/k) \bar{\mathbf{y}}^{k-1}$ and compute $p(\bar{\mathbf{y}}^k)$ as in (7). Get $\boldsymbol{\mu}^k$ as the inner Lagrange multiplier of $p(\bar{\mathbf{y}}^k)$.
- (3) $k = k + 1$.

Until $p(\bar{\mathbf{y}}^k) - d(\bar{\boldsymbol{\mu}}^k) < \epsilon$.

Further details on the MVC decomposition method can be found in [13].

3.2. Coupled-Decompositions Method. Let us consider now the following problem formulation:

$$\begin{aligned}
 \min_{\{\mathbf{x}_j\}, \mathbf{y}} \quad & \sum_{j=1}^J f_j(\mathbf{x}_j) \\
 \text{s.t.} \quad & \mathbf{x}_j \in \mathcal{X}_j, \quad j = 1, \dots, J, \\
 & h_j(\mathbf{x}_j) \leq y_j, \quad j = 1, \dots, J, \\
 & \mathbf{1}^T \mathbf{y} \leq c, \\
 & \mathbf{y} \in \mathcal{Y}, \quad \mathcal{Y} = \mathcal{Y}_1 \times \dots \times \mathcal{Y}_J,
 \end{aligned} \tag{9}$$

where $\mathbf{1}$ is a column vector with all J entries equal to one, and the subset \mathcal{Y} is the cartesian product of J convex one-dimensional subspaces that include the minimum and maximum values of the variables $\{y_j\}$, and thus, it is convex. We consider that $\boldsymbol{\mu}$ is the dual variable associated to the coupling constraint $\mathbf{1}^T \mathbf{y} \leq c$. In the sequel, we briefly describe the algorithm that we propose in order to solve (9). However, the interested reader can find in [27, 28] an extended and well-reasoned version of it.

The technique intertwines the primal and dual subproblems that are obtained when classical primal and dual decompositions [22, Section 6.4] are applied to (9). In primal decomposition, the J subproblems appear when \mathbf{y} is fixed. Note that under this assumption the problem is fully decoupled. Similarly, in dual decomposition we can relax the coupling constraint $\mathbf{1}^T \mathbf{y} \leq c$ (constructing a partial Lagrangian of the problem with dual variable $\boldsymbol{\mu}$), and J subproblems are defined (the problem fully decouples again) for a fixed value of $\boldsymbol{\mu}$. In both classical strategies, the successive updates of \mathbf{y} and $\boldsymbol{\mu}$ are driven by the primal and dual master problems. In the coupled-decompositions method, the result of the primal subproblems is transformed using a redefined dual master problem, the dual projection, and plugged to the dual subproblems. Similarly, the output of the dual subproblems is transformed using the primal projection and fed to the primal subproblems. A flow diagram of the

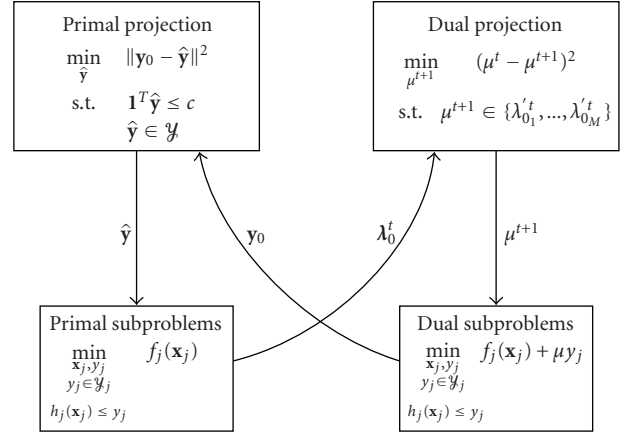


FIGURE 3: Flow diagram of the coupled-decompositions method.

method is depicted in Figure 3. The algorithm starts with $\boldsymbol{\mu}^0 = \mathbf{0}$ and iterates as follows: dual subproblems \rightarrow primal projection \rightarrow primal subproblems \rightarrow dual projection \rightarrow dual subproblems.

Since primal and dual subproblems are extensively analyzed in literature (its formulation appears in Figure 3), let us now detail the novel parts. Notwithstanding, a complete iteration is revisited during the proof of the method. On one hand, primal projection is pretty similar to the primal master problem in primal decomposition. Assuming that \mathbf{y}_0 is constructed with the output of the J dual subproblems, the primal projection solves the following optimization problem:

$$\begin{aligned}
 \min_{\hat{\mathbf{y}}} \quad & \|\mathbf{y}_0 - \hat{\mathbf{y}}\|^2 \\
 \text{s.t.} \quad & \mathbf{1}^T \hat{\mathbf{y}} \leq c, \\
 & \hat{\mathbf{y}} \in \mathcal{Y},
 \end{aligned} \tag{10}$$

with the only particularity that the constraint $\mathbf{1}^T \hat{\mathbf{y}} \leq c$ must be attained with equality when the last update of the Lagrange multiplier is $\boldsymbol{\mu} > \mathbf{0}$. This is in accordance with the Karush-Kuhn-Tucker (KKT) conditions for convex problems [15, Section 5.5] (see more details in [27]). On the other hand, the dual projection takes the output values from the primal subproblems $\boldsymbol{\lambda}_0^t$ and selects the values within $\boldsymbol{\lambda}_0^t$ that have been obtained with primal variables $\hat{\mathbf{y}}_j$ not in the boundary of \mathcal{Y}_j . Let us collect this subset in $\boldsymbol{\lambda}_0^t$. The motivation is that the nonselected values do not directly impact on the value of $\boldsymbol{\mu}$ (it can be seen from the KKT conditions of the problem; see more details in [27]). Thereafter, the $\boldsymbol{\mu}$ update is found as

$$\boldsymbol{\mu}^{t+1} = \arg \left\{ \begin{array}{l} \min_{\boldsymbol{\mu}^{t+1}} (\boldsymbol{\mu}^{t+1} - \boldsymbol{\mu}^t)^2 \\ \text{s.t. } \boldsymbol{\mu}^{t+1} \in \{\boldsymbol{\lambda}_{0_1}^t, \dots, \boldsymbol{\lambda}_{0_M}^t\} \end{array} \right\}, \tag{11}$$

which updates $\boldsymbol{\mu}$ with the value within $\boldsymbol{\lambda}_0^t$ that is closer to the previous $\boldsymbol{\mu}$ value.

Proof of the method: See the appendix.

4. Proposed Solution

Our solution uses a combination of both decomposition techniques. First, an MVC decomposition is applied, making it possible to split the joint problem into one flow or bandwidth allocation subproblem and one subchannel allocation subproblem. The latter depends on variables that are available at the BS, and thus, it is not necessary to explore distributed computations in order to solve it. On the contrary, the former is distributed among the BS and the SSs in order to be standard-compliant (the BS allocates aggregate bandwidth to the SSs and these decide the final allocation to flows and services). In this case, we use a two-level coupled-decompositions strategy.

First, let us consider the problem in (1) and identify rates with \mathbf{x} and subchannel allocation variables with \mathbf{y} in the MVC decomposition formulation in (4). Rewriting the original joint problem as

$$\begin{aligned} \max_{\{r_j^i\}, \{\rho_i\}} \quad & \sum_{i=1}^N \sum_{j=1}^{N_i} U_j^i(r_j^i; p_j^i) \\ \sum_{j=1}^{N_i} r_j^i & \leq \rho_i^T \mathbf{c}_i, \quad i = 1, \dots, N, \\ \{r_j^i\} & \in \mathcal{R}, \\ \{\rho_i\} & \in \mathcal{S}, \end{aligned} \quad (12)$$

where $\mathcal{R} = \{r_j^i \mid m_j^i \leq r_j^i \leq d_j^i\}$ and $\mathcal{S} = \{\rho_i \mid \Gamma \mathbf{1} \preceq \mathbf{1}, \rho_i \succeq \mathbf{0}\}$, we can define the primal subproblem of the MVC decomposition method as

$$\begin{aligned} \max_{\{r_j^i\}} \quad & \sum_{i=1}^N \sum_{j=1}^{N_i} U_j^i(r_j^i; p_j^i) \\ \sum_{j=1}^{N_i} r_j^i & \leq \rho_i^T \mathbf{c}_i, \quad i = 1, \dots, N, \\ \{r_j^i\} & \in \mathcal{R} \end{aligned} \quad (13)$$

for fixed values of $\{\rho_i\}$ and the dual subproblem as

$$\begin{aligned} \max_{\{r_j^i\}, \{\rho_i\}} \quad & \sum_{i=1}^N \sum_{j=1}^{N_i} U_j^i(r_j^i; p_j^i) - \sum_{i=1}^N \gamma_i \left(\sum_{j=1}^{N_i} r_j^i - \rho_i^T \mathbf{c}_i \right), \\ \{r_j^i\} & \in \mathcal{R}, \\ \{\rho_i\} & \in \mathcal{S} \end{aligned} \quad (14)$$

for fixed values of the Lagrange multipliers $\{\gamma_i\}$ that are associated to the constraints that couple rates with subchannel allocation variables in (12). Note that the two subsets of variables are fully decoupled in (14), and thus, the maximization in $\{\rho_i\}$ can be done independently solving the following linear program:

$$\begin{aligned} \max_{\{\rho_i\}} \quad & \sum_{i=1}^N \gamma_i \cdot (\rho_i^T \mathbf{c}_i) \\ \{\rho_i\} & \in \mathcal{S}. \end{aligned} \quad (15)$$

The joint problem is then solved as follows.

Choose a feasible subchannel allocation $\{\rho_i^0\}$ and let $k = 1$.

Repeat

- (1) Let $\rho_i^k = (1/k) \sum_{i=0}^{k-1} \rho_i^{k-1}$, for all i .
- (2) Solve (13) using $\{\rho_i^k\}$ and get the dual variables $\{\gamma_i\}$.
- (3) Let $\gamma_i^k = (1/k) \sum_{i=0}^{k-1} \gamma_i^{k-1}$, for all i .
- (4) Solve (15) using $\{\gamma_i^k\}$ and get updated primal variables $\{\rho_i\}$.
- (5) $k = k + 1$.

Until convergence.

Since (15) is solved at the BS, the remaining issue is to find the solution of (13). In order to avoid excessive DBA-related signaling in the subnetwork and to restrict ourselves to the standard, we propose to solve it using a two-level coupled-decompositions strategy. Note that we can rewrite (13) as

$$\begin{aligned} \max_{\{y^i\}} \quad & \sum_{i=1}^N U^i(y^i) \\ \sum_{i=1}^N y^i & \leq C, \end{aligned} \quad (16)$$

$$\begin{aligned} y^i & \leq \rho_i^T \mathbf{c}_i, \quad i = 1, \dots, N, \\ M^i & \preceq y^i \preceq D^i, \quad i = 1, \dots, N, \end{aligned}$$

where $M^i = \sum_{j=1}^{N_i} m_j^i$, $D^i = \sum_{j=1}^{N_i} d_j^i$, and

$$U^i(y^i) = \begin{cases} \max_{\{r_j^i\}} \sum_{j=1}^{N_i} U_j^i(r_j^i; p_j^i) \\ \text{s.t.} \quad \sum_{j=1}^{N_i} r_j^i \leq y^i, \\ m_j^i \leq r_j^i \leq d_j^i. \end{cases} \quad (17)$$

Note also that the dual Lagrange variable γ_i corresponds to the constraint $y^i \leq \rho_i^T \mathbf{c}_i$ in (16). Therefore, we apply the coupled-decompositions method to solve (16) at the upper layer (BS), and we use it again at the lower layer (at each SS) to solve (17) when it is required by the upper layer.

The iterations of the resulting two-level flow allocation algorithm and the involved signaling are summarized in the following list as well as in Figure 4.

- (1) The dual variable μ^t (associated to $\sum_{i=1}^N y^i \leq C$) is spread through the network, reaching each connection.
- (2) Each connection computes the allocation given μ^t by means of solving the inner dual subproblems (the constraints in m_j^i and d_j^i can be obviated if desired without affecting convergence). The SSs and the BS get their own allocations by aggregation of the allocations below them.

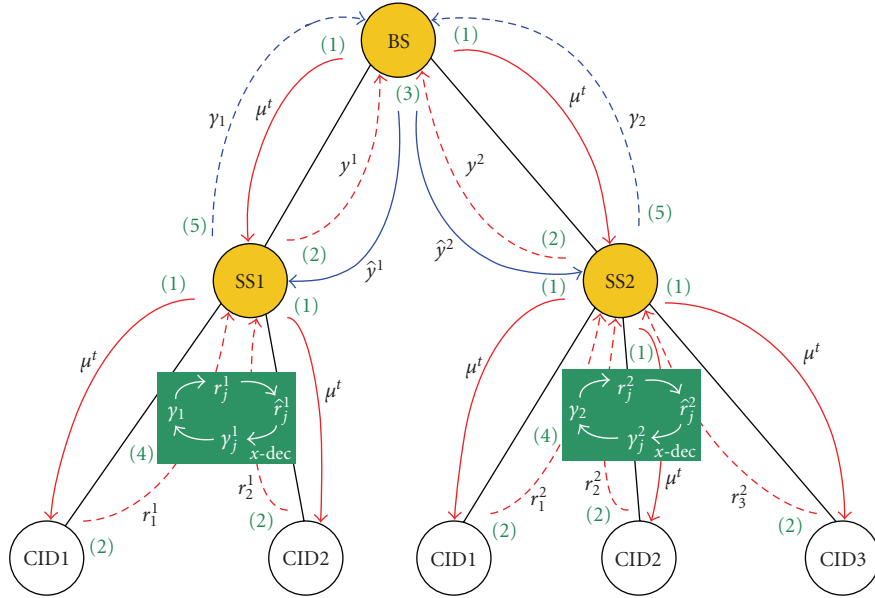


FIGURE 4: 2-level flow allocation algorithm.

- (3) The BS corrects the previous allocations (primal projection) to attain $\sum_{i=1}^N y^i \leq C$ and $y^i \leq \rho_i^T \mathbf{c}_i$, $i = 1, \dots, N$.
- (4) The corrected allocations are used by the SSs to perform inner iterations (within each SS) of the coupled-decompositions method in order to obtain new candidates γ_i .
- (5) Finally, the BS updates the value of the dual variable to μ^{t+1} using the dual projection and the previous γ_i values.

Intuitively, the multilayer coupled-decompositions strategy tries to find a consensus on the price μ that has to be paid for sharing the transport capacity C of the BS. Often, primal variables are interpreted from a resource-oriented perspective whereas dual variables take the role of prices to be paid to use the resources [15, Section 5.4.4]. All CIDs participate in principle in finding such optimal value. However, the price of the connections μ within a particular SS may be distinct from the global price μ if, for example, its link capacity is small (hence forcing the price to locally increase). In these occasions, local prices γ_i that differ from the optimal and global consensus price μ are found.

Other works in literature [10–12] study a similar problem within generic mesh OFDM networks. In general, they search for suboptimal but affordable solutions, which are based on decoupling the joint problem into independent optimization programs that manage only a subset of the variables without looking at the others. In this work, we suggest (for the particular PMP WiMAX case) the derivation of the joint optimal rate and subchannel allocation (under fairness considerations), and we propose a distributed scheme that achieves it. Moreover, the numerical results in the next section show the practical interest of the mechanism in terms of the number of iterations (i.e., directly related to

the amount of signaling). As a matter of fact, the proposed method (possibly with extensions) can be used in other scenarios to speed up the computation of optimization problems or subproblems, either in optimal or suboptimal decoupling approaches.

5. Numerical Results

Let us consider the network setup depicted in Figure 5 with 4 SSs and 9 connections (CIDs) in total. We choose logarithmic utility functions ($\alpha = 1$),

$$U_j^i(r_j^i; p_j^i) = p_j^i \log(r_j^i). \quad (18)$$

Other policies balancing the solution towards the max-sum-rate or the max-min-fair designs can be implemented by fixing other α values using the same algorithm (as discussed later). We fix all requests to 100 kbps (requests are emitted in WiMAX in terms of bytes of information but we transform them to rates taking into account the time basis) and all the minimum granted rates to 1 kbps. All connections have the same priority $p_j^i = 1$. The available number of subchannels is 7, all of them to be shared among the 4 SSs. We consider the following transport capacities (in kbps) per subchannel (10 kHz of bandwidth) and user (given one realization of flat-fading Rayleigh subchannels that have 10 dB of SNR in mean):

$$[\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4] = \begin{bmatrix} 31.49 & 18.58 & 4.07 & 15.69 \\ 34.31 & 13.19 & 29.84 & 24.55 \\ 4.62 & 37.91 & 13.37 & 34.80 \\ 20.54 & 50.62 & 38.91 & 30.92 \\ 34.32 & 22.96 & 27.38 & 48.95 \\ 39.21 & 0.01 & 32.39 & 25.97 \\ 22.10 & 23.69 & 47.14 & 3.86 \end{bmatrix}. \quad (19)$$

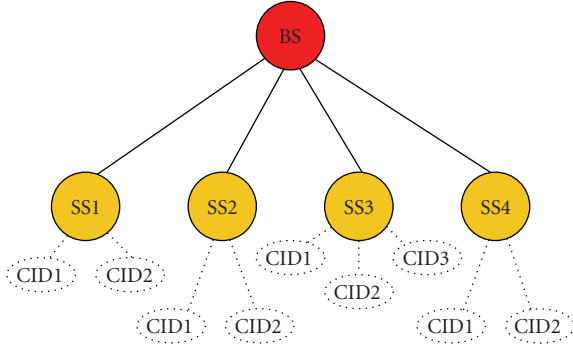
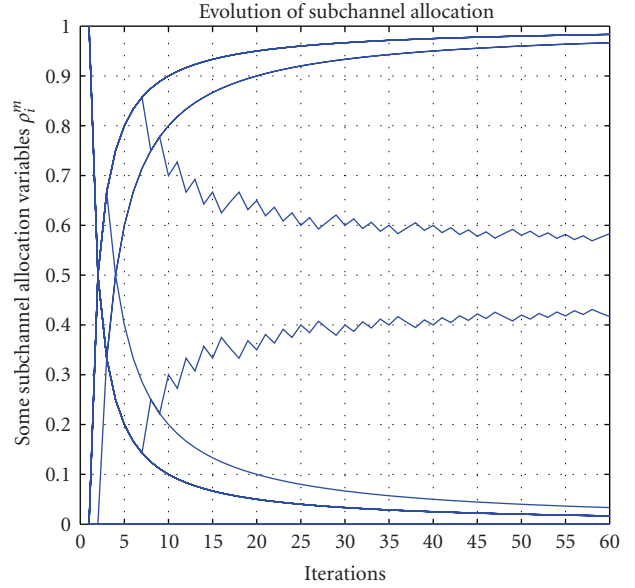


FIGURE 5: Setup of the network under test.

Note that depending on the scheduling length (i.e., the number of contiguous time slots in time that are allocated in a single allocation phase, which fixes the granularity of the ρ_i values) and on the channel characteristics (coherence time), it is reasonable to consider which values of c_i may be really achieved within each allocation phase (mid-term values seem reasonable) so that one may resort to robust designs in order to compute them. The output rate capacity of the BS is 200 kbps, and the initial subchannel allocation is $\Gamma = [\mathbf{I}_{4 \times 4}, \mathbf{0}_{4 \times 3}]^T$ achieving the link capacities $[c^1, c^2, c^3, c^4] = [31.49, 13.19, 13.37, 30.92]$.

Figure 6 shows the evolution of the subchannel allocation variables when we apply the proposed method, achieving new link capacities $[c^1, c^2, c^3, c^4] = [89.39, 86.83, 60.44, 49.23]$. In order to accelerate the convergence to the solution, we have used instantaneous values of $\{\gamma_i\}$ instead of the time-average that is proposed in the MVC decomposition method, averaging only the primal (allocation) variables. This solution has been derived by other authors [8] using a different approach (which validates it), and it is specially relevant in the first iterations where the $\{\gamma_i\}$ values show abrupt changes and very high values. Note that in the figure the final allocation is completely different from the initial one (only SS1 keeps using subchannel 1) but the solution still needs to be rounded to accommodate a practical scheduling implementation, which has its implications also in terms of convergence to the optimal solution because it may have sense to truncate the algorithm after some iterations and round that solution.

In Figure 7, we plot the resulting flow allocation per connection (that correspond to the CIDs ordered from left to right in Figure 5) and the final link capacity once the subchannel allocation has been obtained for the four scenarios specified in Table 1. The objective is to show how fairness considerations impact in the final allocation. The first Scenario is the same as in Figure 6, whereas Scenario 2 evaluates a different allocation scheme (with fairness parameter $\alpha = 0.1$). In the next two scenarios, we study the effect of different priorities using again a proportional fairness approach ($\alpha = 1$). The difference between Scenarios 3 and 4 is that Scenario 3 fixes the same requested rate for all the connections (100 kbps), whereas Scenario 4 has two possible requests (10 kbps and 100 kbps).

FIGURE 6: Evolution of some subchannel allocation variables ρ_i^m .

We notice in the results of Scenario 1 that link capacities have been adjusted (with the subchannel allocation mechanism) in order to provide a similar allocation to all connections. In Scenario 2, the allocation scheme favors the best channels so that each subchannel is assigned to the SS that experiences the maximum achievable rate at that subchannel. Therefore, SS1 gets subchannels 1, 2, and 6; SS2 gets subchannels 3 and 4; SS3 gets subchannel 7; SS4 gets subchannel 5. The corresponding link capacities are $[c^1, c^2, c^3, c^4] = [105.02, 88.54, 47.15, 48.95]$. The final rate allocation is limited by the outcoming rate at the BS (200 kbps) so that Ss 3 and 4 limit their ongoing connections to a lower rate than the connections in Ss 1 and 2, which share the remaining transport capacity. When prioritized traffic flows appear, as in Scenario 3, granted rates are balanced toward services depending on their priority values. Accordingly, it can be seen that subchannel allocation provides more link capacity to Ss 3 and 4. In Scenario 4, we further modify the requested rates with respect to Scenario 3 and the highest priority services in Scenario 3, (the ongoing connections of SS4) reach their requests. As expected, remaining resources (remember that the BS can manage no more than 200 kbps) are redistributed in order to allocate more rate to services in SS3 (with priorities equal to 3) than to services within SS1 and SS2 (with priorities equal to 1), while subchannel allocation favors the link BS-SS3 as well.

Finally, our last result analyzes the efficiency of the novel coupled-decompositions method (used to solve the flow allocation subproblem) in terms of convergence speed. For that purpose, we extend Scenario 1 to 20 SSs with 5 ongoing connections on each. The mean received SNR is 15 dB, and each ongoing connection in Ss 1–15 requests 100 kbps, whereas each connection in Ss 16–20 requests 10 kbps. The transport capacity at the BS is now increased to 1200 kbps.

TABLE 1: Scenario description.

Scenario number	Service priorities p_j^i	Fairness scheme α	Requested rate d_j^i	Granted rate m_j^i
1	All equal to 1	1	All equal to 100 kbps	All equal to 1 kbps
2	All equal to 1	0.1	All equal to 100 kbps	All equal to 1 kbps
3	1 for services in SS1, SS2	1	All equal to 100 kbps	All equal to 1 kbps
	3 for services in SS3			
	5 for services in SS4			
4	1 for services in SS1, SS2	1	100 kbps for services in SS1–SS3	All equal to 1 kbps
	3 for services in SS3		10 kbps for services in SS4	
	5 for services in SS4			

We plot in Figure 8 the evolution of the dual variable μ , that is, negotiated between the BS and the Ss when we use both our novel proposed method and a classic dual decomposition approach using the same 2-layer architecture. Remember that classical decomposition methods need to adjust the value of the step size of the gradient-based update. In this particular case, we have found that a setup with $\alpha(t) = 0.5/t$ at the highest level (i.e., between the BS and the Ss) and $\alpha(t) = 0.005/\sqrt{t}$ at the lowest (i.e., between Ss and connections) provides a satisfactory trade-off between convergence and speed. However, the need of a good adjustment is in practice an obstacle of the method, and it is not easy to find a step providing that good trade-off. On the contrary, one of the important advantages in the coupled-decompositions method is that any user-defined step is completely avoided. The other important advantage is in the number of iterations required. As shown in the figure, the novel technique converges in 5-6 iterations, contrary to the dual decomposition strategy (both obtain the same optimal solution), which needs more than 250 iterations. This drawback of dual decomposition appears in other works in literature, for example, in the numerical results of [10], where it is used to obtain a distributed solution that optimizes power and rate allocation within a mesh OFDM network.

6. Conclusions

In this work, we have proposed an algorithm that implements the DAMA mechanism foreseen in the IEEE 802.16 WiMAX standard. Initially, we have introduced our system model, which considers both flow and subchannel allocations in a cross-layer approach. Some PHY and MAC-layer aspects of WiMAX that are relevant to our work have been briefly reviewed as well as how to translate a series of fairness definitions into a convex optimization framework. All this has led us to formulate a network utility maximization problem.

Since the standard fixes that resources should be requested and granted in a terminal basis but we should consider several traffic flows within each SS (may be with different QoS requirements), we have proposed a distributed solution to the original convex optimization problem in order to fulfill these requirements while keeping the optimality in the allocation. Furthermore, we have explored

the usage of our novel proposed coupled-decompositions algorithm and a recently proposed MVC decomposition method applied to distinct parts of the problem with the goal of achieving a more practical design than with classical primal and dual decompositions.

Results show that it is possible to find a solution to the flow allocation subproblem with very few iterations and without the manual setup of any parameter, as opposite to a classical dual decomposition. The last statement applies also to the subchannel allocation subproblem, which is able to give a good approximation to the solution within a reasonable number of iterations. Finally, we have shown with an example that our strategy is able to attain a fair distribution of resources and to support QoS by means of traffic prioritization.

Appendices

A. Proof of Convergence of the Coupled-Decompositions Method

First of all, we assume that strong duality [15, Section 5.2.3] holds, which is usually verified in convex programs, so that the optimal primal variables attain the optimal dual variables when plugged into the subproblems and vice versa. In the following, the superscript t indicates iteration number although we omit it in some irrelevant occasions. Equivalently, the objective value of the problem is the same regardless it is solved directly (primal version) or by maximizing the dual function (dual version) [15, Section 5.2]. We will prove that

$$\lambda_0^t = \mathbf{1}\mu^t \xrightarrow{t \rightarrow \infty} \lambda^* = \mathbf{1}\mu^*, \tag{A.1}$$

where the relation $\lambda_0^t = \mathbf{1}\mu$ is found by the application of the KKT conditions (see more details in [27]) and μ^* is the optimum value of the dual Lagrange variable. In the following, we review a complete iteration of the method.

Let us consider that $\mu^t < \mu^*$ (the proof is similar if $\mu^t > \mu^*$) and recall the result in [28, Lemma 1], where it is shown that the primal variable \hat{y}_j at the j th subproblem (primal or dual) is a decreasing function of λ_0^t . This fact together with $\lambda_0^t = \mathbf{1}\mu^t$ forces

$$\hat{y}_j(\lambda_0^t) \geq y_j^*, \quad \forall j, \tag{A.2}$$

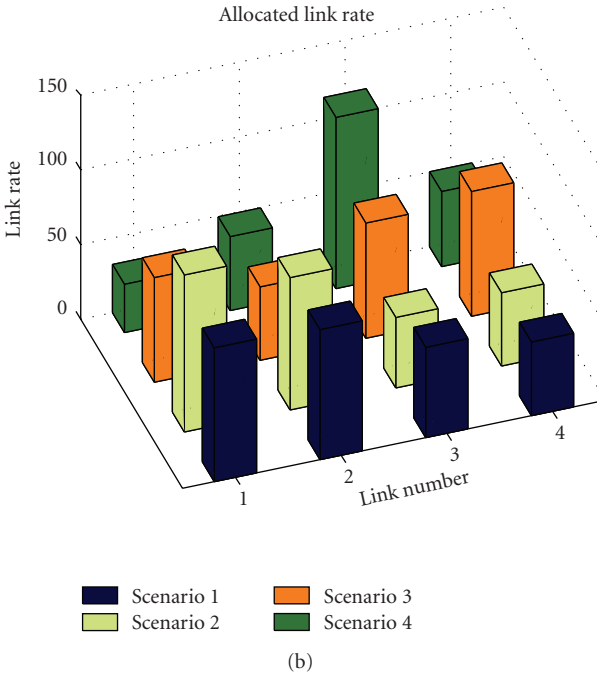
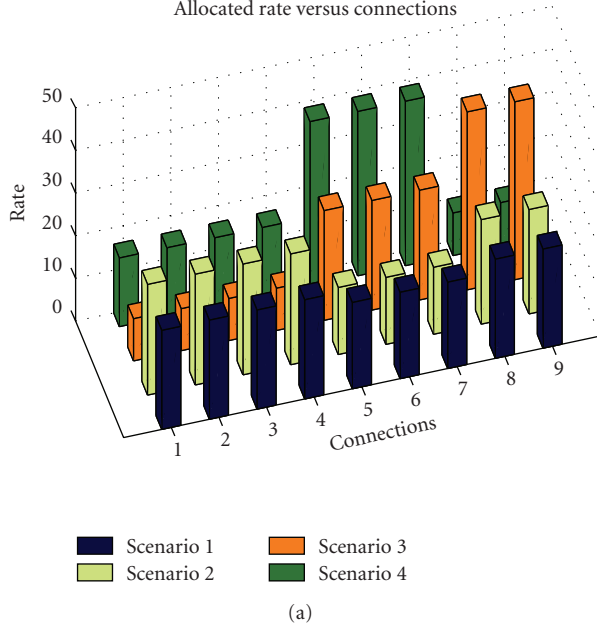


FIGURE 7: Three different allocation examples.

where equality is attained only when $y_j^* \in \text{bd } \mathcal{Y}_j$ (boundary of the subset) and therefore $\mathbf{1}^T \hat{\mathbf{y}} > c$.

In the primal projection, it is verified that

$$\hat{y}_j = y_{0j} - k_j, \quad k_j \geq 0, \quad \forall j \quad (\text{A.3})$$

thanks to the lemma below.

Lemma 1. *Given the optimization problem in (10), its optimal solution can be expressed as $\hat{\mathbf{y}}^* = \mathbf{y}_0 - \mathbf{k}$ with $\mathbf{k} \succeq \mathbf{0}$.*

Proof. See Section B. □

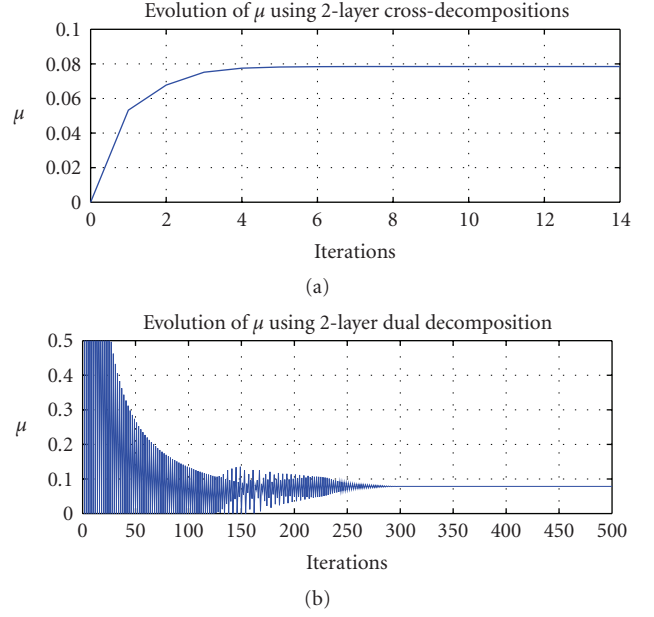
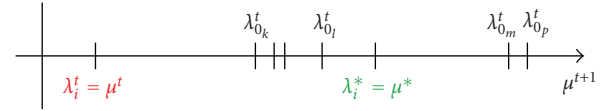
FIGURE 8: Evolution of μ value in the flow allocation subproblem. Comparison between a classical dual decomposition strategy and the proposed coupled-decompositions method.

FIGURE 9: Example of the situation before dual projection.

Applying the relationship between the primal and dual variables of the subproblems to the previous $\hat{\mathbf{y}}^t$ value, it is fulfilled that

$$\lambda_{0j}^t \geq \lambda_j^t, \quad \forall j. \quad (\text{A.4})$$

Furthermore, given that $\hat{\mathbf{y}}^t$ is not the optimal value, it is verified that some of the λ_{0j}^t values are $\lambda_{0j}^t \leq \lambda_j^*$ whereas the remaining ones are $\lambda_{0j}^t \leq \lambda_j^*$, since it holds that $\mathbf{1}^T \hat{\mathbf{y}} = c$. In other words, some of the \hat{y}_j values attain $\hat{y}_j \geq y_j^*$ whereas the rest verify $\hat{y}_j \leq y_j^*$. An example depicting the situation before dual projection can be found in Figure 9.

Consider now that λ_{0j}^t contains a single element. Note that a null vector is not possible since we assume that the coupling constraint is active. Then we can prove the following lemma.

Lemma 2. *Let a primal point $\hat{\mathbf{y}}$ attain $\mathbf{1}^T \hat{\mathbf{y}} = c$ and $\hat{\mathbf{y}} \in \mathcal{Y}$. Let also λ'_0 be a vector containing the dual translation (computed by primal subproblems) of the values in $\hat{\mathbf{y}}$ that verify $\hat{\mathbf{y}} \in \text{int } \mathcal{Y}$ (interior of the subset). Then, if the vector λ'_0 is in fact a scalar, it is verified that*

$$\lambda'_0 \leq \lambda'^* = \mu^*, \quad (\text{A.5})$$

where λ'^* is the optimum value of λ for the selected position in λ'_0 (i.e., equal to μ^*).

Proof. Using Lemma 1, we can state that all the values within $\hat{\mathbf{y}}$ except the k th element accomplish $\hat{y}_i \in \inf \mathcal{Y}_i$ ($i \neq k$). Therefore, it holds that $\hat{y}_k < \hat{y}_k^* = y_{0k}^* = y_k^*$. Applying the relationship between subproblems (remember that both in primal and dual subproblems, primal variables are a decreasing function of dual variables and $\hat{\mathbf{y}}(\lambda_0^*) = \mathbf{y}^*$), we reach the desired result. \square

Finally, we update μ^{t+1} using (11). Collecting all the results obtained up to this point, we have that

$$\mu^{t+1} > \mu^t \quad (\text{A.6})$$

since every value in λ_0^t verifies $\lambda_{0i}^t > \mu^t$. Furthermore, it is also true that

$$\mu^{t+1} < \mu^* \quad (\text{A.7})$$

since the value λ_{0i}^t closer to μ^t (dual projection) accomplishes $\lambda_{0i}^t < \lambda_i^* = \mu^*$, which is derived from Lemma 2 and the discussion preceding it. Figure 9 provides a graphical explanation. We can finally conclude that

$$\mu^t < \mu^{t+1} < \mu^*. \quad (\text{A.8})$$

The proof ends showing by contradiction that μ^t cannot tend to a value smaller than μ^* . Assume that there exists a value μ^\triangleright where successive iterations converge. Then μ^\triangleright is a stationary point of the method. In other words, a complete iteration of the method starting from μ^\triangleright returns exactly the same value. This enforces in the primal projection that $\hat{\mathbf{y}} = \mathbf{y}_0(\mu^\triangleright)$, otherwise the values in λ_0^t would increase and so the update in μ (dual projection). Given the relationship between primal and dual subproblems, we see that the previous equation is only attained if $\mu^\triangleright = \mu^*$ since a lower value $\mu^\triangleright < \mu^*$ would obtain a primal point $\mathbf{y}_0(\mu^\triangleright)$ from dual subproblems such that $\mathbf{1}^T \mathbf{y}_0(\mu^\triangleright) > c$.

Before concluding this section, we want to note that it is possible to substitute the primal projection by the projection into $\mathbf{1}^T \mathbf{y} = c$ and the method still converges (it can be similarly proved). It is a more practical option since the projection can be analytically computed as [15, Section 8.1]

$$\hat{\mathbf{y}}^t = \mathbf{y}_0^t + \frac{(c - \mathbf{1}^T \mathbf{y}_0^t) \mathbf{1}}{J}. \quad (\text{A.9})$$

B. Proof of Lemma 1

First, note that a point $\hat{\mathbf{y}} = \mathbf{y}_0 - \mathbf{k}$ with $\mathbf{k} \succcurlyeq \mathbf{0}$ is feasible since it attains both $\mathbf{1}^T \hat{\mathbf{y}} \leq c$ and $\hat{\mathbf{y}} \in \mathcal{Y}$ (assuming that the intersection is not empty). Then, we have to prove that a point that does not accomplish the equation $\hat{\mathbf{y}} = \mathbf{y}_0 - \mathbf{k}$ for positive values in \mathbf{k} cannot be optimal for problem (10).

We proof this last result by induction. Assume a certain vector \mathbf{k} , called $\mathbf{k}^\triangleright$ that attains $\mathbf{1}^T (\mathbf{y}_0 - \mathbf{k}^\triangleright) = c$ and $\mathbf{k}^\triangleright \succcurlyeq \mathbf{0}$. Construct now a new vector \mathbf{k}^\dagger from $\mathbf{k}^\triangleright$ by fixing its l th element k_l^\dagger to $-a$ with $a > 0$ and distributing the difference $|k_l^\triangleright - k_l^\dagger|$ among the rest of elements in \mathbf{k}^\dagger so as to attain the

equality coupling constraint. In other words,

$$k_i^\dagger = \begin{cases} -a, & i = l \\ k_i^\triangleright + \epsilon_i, & i \neq l, \epsilon_i > 0 \end{cases}, \quad \sum_i k_i^\dagger = \mathbf{1}^T \mathbf{y}_0 - c. \quad (\text{B.10})$$

Let us introduce some results from majorization theory [29] that we need to complete the proof. First, let the components of $\mathbf{x} \in \mathbb{R}^n$ be ordered in decreasing order and express it as

$$x_{[1]} \geq \dots \geq x_{[n]}. \quad (\text{B.11})$$

Then, it is said [29, 1.A.1] that a vector \mathbf{y} majorizes a vector \mathbf{x} (which we denote by $\mathbf{y} \succ^M \mathbf{x}$), $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ if

$$\begin{aligned} \sum_{i=1}^k x_{[i]} &\leq \sum_{i=1}^k y_{[i]}, \quad k = 1, \dots, n-1, \\ \sum_{i=1}^n x_{[i]} &= \sum_{i=1}^n y_{[i]}. \end{aligned} \quad (\text{B.12})$$

From the definition above and the construction process of \mathbf{k}^\dagger , we can state that $\mathbf{k}^\dagger \succ^M \mathbf{k}^\triangleright$.

Second, a real-valued function ϕ on a set $\mathcal{A} \subseteq \mathbb{R}^n$ is called Schur-convex if [29, 3.A.1]

$$\mathbf{y} \succ^M \mathbf{x} \text{ on } \mathcal{A} \implies \phi(\mathbf{y}) \geq \phi(\mathbf{x}). \quad (\text{B.13})$$

And third, a function $\phi(\mathbf{x}) = \sum_i g(x_i)$, where g is convex, is Schur-convex [30, Corollary 3.1].

With those results in hand, we want to compare $\|\mathbf{y}_0 - \hat{\mathbf{y}}\|^2$ for $\mathbf{k} = \mathbf{k}^\triangleright$ and $\mathbf{k} = \mathbf{k}^\dagger$. Let us rewrite the quadratic norm as

$$\|\mathbf{y}_0 - \hat{\mathbf{y}}\|^2 = \|\mathbf{y}_0 - \mathbf{y}_0 + \mathbf{k}\|^2 = \sum_i k_i^2 \quad (\text{B.14})$$

and consider $\phi(\mathbf{k}) = \sum k_i^2$, which is a Schur-convex function. Finally, since $\mathbf{k}^\dagger \succ^M \mathbf{k}^\triangleright$, we have

$$\|\mathbf{k}^\dagger\|^2 \geq \|\mathbf{k}^\triangleright\|^2, \quad (\text{B.15})$$

and thus, any solution where one element within \mathbf{k} is negative is not optimal (since the problem is convex and has a single solution). The proof ends by induction of this result to an arbitrary number of negative elements in \mathbf{k} .

Notation

$U_i(r_i; p_i, \alpha)$:	Utility achieved when entity i transmits at rate r_i . The utility is parameterized by a priority p_i (entity-dependant) and a shape factor α (common to all utilities)
N :	Number of SSs
N_i :	Number of active connections at the i th SS
r_j^i :	Rate of the j th ongoing connection at the i th SS
m_j^i :	Minimum guaranteed rate to the j th ongoing connection at the i th SS

- d_j^i : Requested rate of the j th ongoing connection at the i th SS
- C : Maximum outgoing rate at the BS
- ρ_i : Subchannel allocation vector at the i th SS
- c_i : Achievable rates at the i th SS (includes all subchannels)
- $c^i(\rho_i)$: Maximum outgoing rate at the i th SS
- Γ : Subchannel allocation matrix:
 $\Gamma = [\rho_1, \dots, \rho_N]$
- \mathcal{R} : Feasible rates subset: $\mathcal{R} = \{r_j^i | m_j^i \leq r_j^i \leq d_j^i\}$
- \mathcal{S} : Feasible allocations subset:
 $\mathcal{S} = \{\rho_i | \Gamma \mathbf{1} \preceq \mathbf{1}, \rho_i \succeq \mathbf{0}\}$

Acknowledgments

This work was supported in part by the Spanish Ministry of Science and Innovation under TEC200806305 project.

References

- [1] IEEE, "Air Interface for Fixed Broadband Wireless Access Systems," IEEE Standards, October 2004.
- [2] IEEE, "Air Interface for Fixed and Mobile Broadband Wireless Access Systems; Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Band and Corrigendum 1," IEEE Standards, February 2006.
- [3] ETSI, "Broadband Radio Access Networks (BRAN); HIPERMAN; Data Link Control (DLC) Layer," ETSI TS 102 178, March 2003.
- [4] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*, Prentice-Hall, Englewood Cliffs, NJ, USA, 2007.
- [5] B. Makarevitch, "Adaptive resource allocation for WiMAX," in *Proceedings of the 18th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '07)*, pp. 1–6, Athens, Greece, September 2007.
- [6] H.-Y. Wei, S. Ganguly, R. Izmailov, and Z. J. Haas, "Interference-aware IEEE 802.16 WiMax mesh networks," in *Proceedings of the 61st IEEE Vehicular Technology Conference (VTC '05)*, vol. 5, pp. 3102–3106, Stockholm, Sweden, May 2005.
- [7] P. Du, W. Jia, L. Huang, and W. Lu, "Centralized scheduling and channel assignment in multi-channel single-transceiver WiMax mesh network," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC '07)*, pp. 1736–1741, Hong Kong, March 2007.
- [8] P. Soldati, B. Johansson, and M. Johansson, "Distributed optimization of end-to-end rates and radio resources in WiMax single-carrier networks," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '06)*, pp. 1–6, San Francisco, Calif, USA, November 2006.
- [9] P. Soldati and M. Johansson, "Network-wide resource optimization of wireless OFDMA mesh networks with multiple radios," in *Proceedings of IEEE International Conference on Communications (ICC '07)*, pp. 4979–4984, Glasgow, UK, June 2007.
- [10] L. B. Le and E. Hossain, "Joint rate control and resource allocation in OFDMA wireless mesh networks," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC '07)*, pp. 3041–3045, Hong Kong, March 2007.
- [11] K.-D. Lee and V. C. M. Leung, "Fair allocation of subcarrier and power in an OFDMA wireless mesh network," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 11, pp. 2051–2060, 2006.
- [12] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2726–2737, 2005.
- [13] K. Holmberg and K. C. Kiwiel, "Mean value cross decomposition for nonlinear convex problems," *Optimization Methods and Software*, vol. 21, no. 3, pp. 401–417, 2006.
- [14] L. Reggiani, L. G. Giordano, and L. Dossi, "Multi-user subchannel, bit and power allocation in IEEE 802.16 systems," in *Proceedings of the 65th IEEE Vehicular Technology Conference (VTC '07)*, pp. 3120–3124, Dublin, Ireland, April 2007.
- [15] L. Boyd and S. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2003.
- [16] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 10, pp. 1747–1758, 1999.
- [17] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 556–567, 2000.
- [18] D. Bertsekas and R. Gallager, *Data Networks*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1987.
- [19] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, no. 1, pp. 33–37, 1997.
- [20] A. Muthoo, *Bargaining Theory with Applications*, Cambridge University Press, Cambridge, UK, 1999.
- [21] H. Yaïche, R. R. Mazumdar, and C. Rosenberg, "A game theoretic framework for bandwidth allocation and pricing in broadband networks," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 667–678, 2000.
- [22] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, Mass, USA, 1999.
- [23] L. S. Lasdon, *Optimization Theory for Large Systems*, Dover, New York, NY, USA, 2002.
- [24] S. H. Low and D. E. Lapsley, "Optimization flow control—I: basic algorithm and convergence," *IEEE/ACM Transactions on Networking*, vol. 7, no. 6, pp. 861–874, 1999.
- [25] J.-W. Lee, M. Chiang, and A. R. Calderbank, "Network utility maximization and price-based distributed algorithms for rate-reliability tradeoff," in *Proceedings of the 25th IEEE International Conference on Computer Communications (INFOCOM '06)*, pp. 1–13, Barcelona, Spain, April 2006.
- [26] D. P. Palomar and M. Chiang, "Alternative distributed algorithms for network utility maximization: framework and applications," *IEEE Transactions on Automatic Control*, vol. 52, no. 12, pp. 2254–2269, 2007.
- [27] A. Morell, G. Seco-Granados, and M. A. Vázquez-Castro, "Computationally efficient cross-layer algorithm for fair dynamic bandwidth allocation," in *Proceedings of the 16th International Conference on Computer Communications and Networks (ICCCN '07)*, pp. 13–18, Honolulu, Hawaii, USA, August 2007.
- [28] A. Morell, G. Seco-Granados, and J. L. Vicario, "Distributed algorithm for uplink scheduling in WiMAX networks," in *Proceedings of the 5th International Conference on Broadband*

Communications, Networks, and Systems (BROADNETS '08), pp. 257–264, London, UK, September 2008.

- [29] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York, NY, USA, 1979.
- [30] D. P. Palomar, *A unified framework for communications through MIMO channels*, Ph.D. dissertation, Technical University of Catalonia (UPC), Barcelona, Spain, May 2003.