Using Lagrangian Relaxation for Radio Resource Allocation in High Altitude Platforms

Ahmed Ibrahim and Attahiru S. Alfa

***Abstract*—In this paper, we study radio resource allocation for multicasting in OFDMA based high altitude platforms (HAPs). We formulate and solve an optimization problem that finds the best allocation of HAP resources such as radio power, subchannels, and time slots. The problem also finds the best possible frequency reuse across the cells that constitute the service area of the HAP. The objective is to maximize the number of user terminals that receive the requested multicast streams in the HAP service area in a given OFDMA frame. A bounding subroutine in a branch and bound algorithm can be obtained by decomposing it into two easier subproblems, due to its high complexity, and solving them iteratively. Subproblem 1 turns out to be a binary integer linear program of no explicitly noticeable structure and therefore Lagrangian relaxation is used to dualize some constraints to get a structure that is easy to solve. Subproblem 2 turns out to be a linear program with a continuous knapsack problem structure. Hence a greedy algorithm is proposed to solve subproblem 2 to optimality. The subgradient method is used to solve for the dual variables in the dual problem to get the tightest bounds.**

***Index Terms*—High altitude platforms, multicasting, radio resource allocation, continuous knapsack problem, lagrangian relaxation.**

I. Introduction

High altitude platforms (HAPs) are quasi-stationary aerial platforms that are meant to be located at a height of 17–22 km above Earth's surface in the stratosphere layer. Many of their pros are a combination of those in both, terrestrial wireless and satellite communication systems. Some of those pros are [1]:

- Their ability to fly on demand to temporarily or permanently serve regions with unavailable telecommunications infrastructure.
- A single HAP has a large area coverage compared to a single terrestrial cellular base station (BS) with a radius that can go up to 150 km.
- Low propagation delays compared to satellites which implies better perceived quality of service (QoS) by the users for real time applications like voice and video.
- Stronger received signal strengths as compared to satellites and hence user terminals need not be bulky.
- Deployment time is low since one platform and ground support are sufficient to start the service.
- Much less ground-based infrastructure compared to terrestrial cellular networks.

The amount of published research in RRA over HAPs is not much, inspite of its importance and the challenge it poses, and almost all consider UMTS-CDMA wireless interface which differs a lot from OFDMA based interface. Some of the published work in HAP RRA came in [2]–[9]. In this paper, we consider radio resource allocation (RRA) for OFDMA based HAP system with multicasting in the downlink in order to best utilize the scarce radio resources to maximize the number of user terminals (UTs) that can be admitted to receive the requested multicast session in a given OFDMA frame. We exploit an important feature that HAPs have over terrestrial wireless communications networks that is central power management at the platform level rather than at the base station (HAP antenna) level [5]. This yields a greater flexibility in power allocation. Furthermore, we also use central management for all the frequency-time slots of the OFDMA frame at the HAP level and dynamically reuse them across the cells that constitute the HAP service area provided a signal-to-interference-noise-ratio (SINR) threshold is satisfied.

An optimization problem is formulated with the objective of maximizing the total number of UTs receiving transmission in a given OFDMA frame while satisfying certain quality of service (QoS) requirements like bit-error rate (BER) and data rate with the available resources. The problem turns out to be a complicated *Mixed Integer Non-Linear Program* (MINLP) [10] which we decompose to two easier subproblems and solve each separately to get an acceptable solution bound.

Our proposed technique requires solving the subproblems sequentially using the solution of one to solve the other and to iterate back and forth between the two subproblems as in [8] to obtain a better solution. The first subproblem turns out to be a *Binary Integer Linear Program* (BILP) whose solution decides which UTs to select for service in a multicast session group in a particular cell and which frequency-time slots to allocate to the

This research is supported in part by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

A. Ibrahim is with the Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB R3T 5V6, Canada (e-mail: Ahmed.Ibrahim@umanitoba.ca).

A. S. Alfa is with the Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB R3T 5V6, Canada, and also with the Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa (e-mail: Attahiru.Alfa@umanitoba.ca).

chosen user terminals. The second subproblem is a power allocation subproblem. It is a linear program (LP) [11] problem that allocates power for different multicast sessions across different cells in the HAP service area in each frequency-time slot.

For subproblem 1, we propose different methods based on Lagrangian relaxation [12], [13] and use the subgradient method [14] to find the optimal dual variable values of the subproblem. We compare between those different methods in terms of complexity and bound goodness. For subproblem 2, we were able to model the problem as a continuous knapsack problem (CKS) and proposed a greedy approach to solve it optimally. The obtained solution is an upper bound on the optimal solution to subproblem 1 that can be used as a bounding technique in *branch and bound* (BnB) algorithm [11] to get a feasible solution to the subproblem.

II. Related Work on RRA for Multicasting in OFDM Systems

While lots of research exists on RRA in unicast terrestrial multiuser OFDM systems, research on multicast RRA is still emerging in broadband wireless systems in general. There are two types of multicast transmissions, single-rate transmissions and multi-rate transmissions. In single-rate, the transmission to all users in each multicast group is with the same rate irrespective of their non-uniform achievable capacities. In multi-rate however, the transmission to each user in each multicast group can be at different rates based on what each user can handle.

There are three simple schemes that have been adopted for single-rate transmissions [15]:

1) Predefined fixed default group rates for the entire multicast group. CDMA 2000 1x EV-DO networks for example, use 204.8 kbps for multicast transmission and assign resources equally to all users in cyclic round-robin.
2) *Least Channel Gain (LCG) User Rate*: This scheme adaptively sets the group transmission rate according to the user with the poorest channel conditions.
3) Average (AVG) Group Throughput: The transmission rate to each group is based on the long-term moving average throughput of the group. There are various techniques for the group averaging, for e.g. in [16] the median throughput is selected among the achievable instantaneous throughput of the users in the group such that half of the members can be supported. Also, in [17] the single transmission rate is selected based on the exponential moving average received throughput of each user in the cell.

For multi-rate multicast transmissions two techniques currently exist in the literature [15]:

1) *Information Decomposition Techniques* (IDT): High rate multimedia content is split into multiple sub-streams where users can subscribe to the amount of data that can reliably be received. It exploits user multichannel diversity and the perceived quality improves as the user receives more substreams. There are two types of IDT available in literature: Multiple Description Coding (MDC) and Hierarchical Layering (HL).
2) *Multicast Subgroup Formation* (MSF): This involves splitting multicast groups into smaller subgroups based on the intra-group users' channel qualities and single multicast transmission rate is defined for each subgroup. It combines the simplicity of single rate and higher capacity of IDT.

Single rate is popular for its implementation simplicity and low complexity. This scheme guarantees reliable multicasting to user terminals with poor link conditions. However it puts severe restriction on the achievable throughput. Multi-rate, on the other hand, has been receiving more attention lately because of the necessity to achieve user throughput differentiation such that improved system spectral efficiency is attained. Multi-rate multicasting addresses the sub-optimality that exists in single-rate transmission considering the intrinsic heterogeneous channel characteristics of each user terminal in a multicast group. Its aim is to alleviate the problem of intra-group unfairness in single-rate transmissions caused by large differences in channel conditions for different users in a multicast group. However it exhibits higher coding and synchronization complexities, requires high computational power and requires heavy retransmission overhead [15].

In the literature, there are three main categories for multicasting RRA, those are *Strict Throughput Maximization* (STM), *Max-Min Fairness* (MMF) and *Proportional Throughput with Fairness* (PTF). For each, various algorithms and optimization techniques have been proposed, those are summarized in [15].

In this paper, we choose single rate transmissions for its advantages that were mentioned earlier, especially that our resource allocation problem is a very complex one with five degrees of freedom. Hence the low complexity feature in single rate transmission is desirable. Also, in our problem we are trying to maximize the number of served (or admitted) users in the service area rather than maximizing the throughput. So the drawback of the LCG single rate scheme mentioned earlier is not a concern for us.

Existing studies in radio resource allocation for multicast services in OFDMA systems have mainly investigated a single multicast session. Those studies, along with the few that considered multiple multicast sessions, mostly considered a single cell system [15]. As an effort to fill in this gap, we combine multiple muticast sessions with a multicellular HAP system.

III. System Model

We consider the scenario of a single HAP at an altitude of 20 km in the stratosphere providing cellular coverage to a number of UTs in a given geographical area. Unlike terrestrial cellular systems, all the cells in the service area have one common platform power source with a total power P_{PF}^{total} which allows for better and flexible power utilization. Also, the HAP centrally manages all the radio resources (power, subchannels and time slots) assignment among all the cells in its service area which is quite complicated. This is not the case usually in terrestrial cellular networks which have many base stations and different network entities where distribution and cooperation can be used to make the resource allocation problem simpler than that in HAP.

The OFDMA frame consists of a fixed number of subchannels C and an equal number of time slots T on each subchannel. Therefore, a frequency-time slot refers to one subchannel c in time slot t, i.e. a slot (c, t), with ΔB as its bandwidth and ΔT its time duration. The gains $g_{i,k,c,t}$ for each slot (c, t) for a UT k in cell i are independent and we assume that their values are known at the HAP side. The channel gains depend upon the instantaneous values of large scale fading and small scale fading. In a HAP system, large scale fading is a result of free space path loss and attenuation due to rain and clouds [18]. Small scale fading is acceptably modeled as Ricean fading due to the presence of line of sight rays from the HAP to most of the locations in the HAP service area [19]. The channel gain $g_{i,k,c,t}$ between base station (antenna) i and UT k on the frequency-time slot (c, t) can hence be given as:

$$g_{i,k,c,t} = \left(\frac{C_{light}}{4\pi d_k f_c}\right)^2 .G_H(\Psi_{i,k}).G_k^u.A(d_k).\varphi_{k,c,t} \tag{1}$$

where

- $G_H(\Psi_{i,k})$ is the gain seen at an angle $\Psi_{i,k}$ between user terminal k and antenna i boresight axis.
- d_k is the distance between the HAP and user terminal k.
- C_{light} is the speed of light and f_c is the carrier frequency.
- $A(d_k)$ is the attenuation due to clouds and rain. This depends on the distance between the HAP and each user k in the service area.
- G_k^u the antenna's gain of user terminal k.
- $\varphi_{k,c,t}$ is the Ricean small scale gain in frequency-time slot (c, t) for user terminal k.

All the antennas are co-located on the HAP. The interference to a user in a particular cell is due to the reception of unwanted transmissions at boresight angles greater than angles that subtend the neighbor cell footprints through the main-lobes and side lobes of their antennas [20]. A frequency-time slot is reused across different cells for any multicast sessions as long as an SINR threshold γ^{th} for the UTs satisfy an acceptable BER in that slot. However within any cell i, a frequency-time slot cannot be assigned to more than one multicast session m.

Let the set of UTs that get admitted to multicast session m in cell i for a given OFDMA frame be $N_{m,i}$. The transmission rate $r_{m,i,c,t}$ in a frequency-time slot (c, t) in cell i for session m has to be supported by the UTs with the worst SINR among all the UTs $k \in N_{m,i}$ to ensure they can all properly receive the multicast session. We define $r_{m,i,c,t}$ as Shannon's capacity on slot (c, t) for the UT with the poorest link conditions in $N_{m,i}$:

$$r_{m,i,c,t} = \frac{\Delta B \Delta T}{F} \log_2\left(1 + \min_{k \in N_{m,i}} \gamma_k^{c,t}\right), \tag{2}$$

Table I
NOTATION DEFINITIONS

Notation	Definition
$\mathcal{S}=\{1,2,...,S\}$	set of cells in the HAP service area.
$\mathcal{M}=\{1,2,...,M\}$	set of sessions being multicasted in the HAP service area.
$\mathcal{C}=\{1,2,...,C\}$	set of subchannels available for the HAP service area
$\mathcal{T}=\{1,2,...,T\}$	set of time slot indices on a given subchannel c
$\mathcal{K}=\{1,2,...,K\}$	set of UT indices in HAP service area
$s_{m,i} \subseteq \mathcal{K}$	set of UTs that request session m in cell i
$\lambda_{m,i,k}$	a binary constant indicating whether UT k is tuned to receive session m in cell i
$z_{m,i,k,c,t}$	is a binary variable indicating whether UT k receives a transmission for session m in cell i in a given OFDMA frame slot (c, t).
$p_{m,i,c,t}$	is a non-negative variable indicating the power level in slot (c, t) of cell i for session m.
$x_{m,i,c,t}$	is a binary variable that indicates whether slot (c, t) is assigned to session m in cell i.

where F is the OFDMA frame duration and $\gamma_k^{c,t}$ is the SINR of UT k on slot (c, t) and is defined as

$$\gamma_k^{c,t} = \frac{g_{i,k,c,t}p_{m,i,c,t}}{\sum_{m=1}^{M} \sum_{\forall i' \in \mathcal{S}, i' \neq i} g_{i',k,c,t}p_{m,i',c,t} + \sigma^2}, \tag{3}$$

where $\mathcal{S}$ is the set of cell indices, $p_{m,i,c,t}$ is the power allocated for session m in cell i over slot (c, t) and σ^2 is the noise power per subchannel.

In each cell i in the HAP service area, there is a set of UTs that are tuned to receive (or request to receive) a multicast session m, let this set be $s_{m,i}$. We are trying to find the following such that the number of UTs receiving transmission in a given OFDMA frame is maximum:

- Which UTs from $s_{m,i}$ to assign to the set of user terminals $N_{m,i} \subseteq s_{m,i}$.
- Which frequency-time slots to assign to those users in $N_{m,i}$ for all sessions in all cells.
- The power level for each multicast session m in each cell i in each frequency-time slot (c, t).

The quality of service (QoS) requirements which the RRA must satisfy are:

- Minimum SINR requirement for each UT k receiving a multicast transmission m in a cell i over a slot (c, t) i.e. $\gamma_k^{c,t} \geq \gamma^{th}$.
- Minimum and maximum capacity requirements (R_m and C_m) for multicast session m in cell i given by $R_m \leq \sum_{c=1}^{C} \sum_{t=1}^{T} \frac{\Delta B \Delta T}{F} \log_2\left(1 + \min_{k \in N_{m,i}} \gamma_k^{c,t}\right) \leq C_m$.

The optimization problem is therefore a joint admission control-RRA-scheduling scheme to maximize the number of UTs receiving multicast transmissions in the entire HAP service area over a given OFDMA frame.

IV. Problem Formulation

In this section, we illustrate our optimization problem for the RRA. Table I gives the definitions of the different variables, constants and notations. The optimization problem can be given as:

Problem O :

$$\max_{z_{m,i,k,c,t},p_{m,i,c,t},x_{m,i,c,t}} \sum_{m=1}^{M}\sum_{i=1}^{S}\sum_{k=1}^{K}\sum_{c=1}^{C}\sum_{t=1}^{T} z_{m,i,k,c,t} \tag{O}$$

s.t.

$$C1 : z_{m,i,k,c,t} = \begin{cases} \{0,1\} & k : \lambda_{m,i,k} = 1, \\ 0 & \text{otherwise}, \end{cases} \qquad \forall m,i,k,c,t$$

$$C2 : z_{m,i,k,c,t,} = z_{m,i,k,c',t'}, \qquad \forall c,t : x_{m,i,c,t} = 1, \quad \forall c',t' : x_{m,i,c',t'} = 1, \forall m,i,k$$

$$C3 : \gamma_k^{c,t} \geq \gamma^{th} \qquad \forall k : z_{m,i,k,c,t} \neq 0, \qquad \forall m,i,c,t$$

$$C4 : R_m \leq \sum_{c=1}^{C}\sum_{t=1}^{T} r_{m,i,c,t} \leq C_m \qquad \forall k : z_{m,i,k,c,t} \neq 0, \forall m,i$$

$$C5 : x_{m,i,c,t} = \begin{cases} 1 & > 0, \\ 0 & \text{otherwise}, \end{cases} \qquad \forall m,i,c,t$$

$$C6 : \sum_{m=1}^{M} x_{m,i,c,t} \leq 1, \qquad \forall i,c,t$$

$$C7 : \sum_{m=1}^{M}\sum_{i=1}^{S}\sum_{c=1}^{C} \leq P_{PF}^{total}, \qquad \forall t$$

$$C8 : p_{m,i,c,t} \geq 0 \qquad \forall m,i,c,t.$$

The following is a brief explanation of the objective function and the constraints:

- We construct a single aggregate objective function by summing the user terminals in all frequency-time slots in all cells and multicast sessions.
- Constraint $C1$ means that a user terminal k may be assigned to receive a multicast transmission for session m in cell i in an OFDMA frame slot (c,t) only if it is tuned to (or requests) that session and resides in the cell i.
- Constraint $C2$ is to ensure that user terminals assigned to receive session m in cell i over a set of slots in the OFDMA frame should be the same in each of those assigned slots since all users receiving multicast transmission share the same resources.
- In constraint $C3$ the selection of user k residing in cell i to receive a session m in a particular slot (c,t) (i.e. the value of $z_{m,i,k,c,t}$) depends on whether there is enough power for that user to join the multicast group m and receive transmission in slot (c,t) of cell i while satisfying the minimum SINR value. This is important since different users have different channel gains on each slot (c,t).
- In constraint $C4$ the value of $z_{m,i,k,c,t}$ depends on whether there is enough power for that user to satisfy the minimum and maximum rate constraints.
- Constraint $C5$ defines a binary integer variable $x_{m,i,c,t}$ whose value is 1 if $p_{m,i,c,t}$ is non-zero.
- Constraint $C6$ implies that $p_{m,i,c,t}$ can only be non zero for at most one session in a particular slot (c,t) for a cell i.
- Constraint $C7$ guarantees that for a given time slot, the total power of the HAP is not exceeded and $C8$ guarantees that the power values are always non-negative.

Using some algebraic and logical manipulation, we were able to obtain an easier formulation with the variables $z_{m,i,k,c,t}$ and $p_{m,i,c,t}$ only. The reformulation is obtained by: 1) replacing the slot indicator variable $x_{m,i,c,t}$ in **Problem O** and its corresponding constraints $C6$ with a set of linear constraints in $z_{m,i,k,c,t}$ and $p_{m,i,c,t}$ and 2) getting around the difficulty with the capacity constraint $C4$ by replacing it with non-logarithmic constraint sets in $z_{m,i,k,c,t}$ and $p_{m,i,c,t}$. Either the lower capacity bound constraint or the SINR constraint can be selected and the other can be ignored due to its redundancy. The following constraints in **Problem O** maybe rewritten as follows:

- $C1$ can be written as:

$$0 \leq z_{m,i,k,c,t} \leq \lambda_{m,i,k}, \qquad \forall m,i,k,c,t$$

- $C2$ can be written as:

$$z_{m,i,k,c,t} + z_{m,i,k',c',t'} \leq 1 + z_{m,i,k,c',t'},$$
$$\forall m,i; \forall k,k' : k \neq k'; \forall c,c' : c \neq c'; \forall t,t' : t \neq t'$$

- $C5$ and $C6$ can be replaced by:

$$z_{m',i,k',c,t} \leq 1 - z_{m,i,k,c,t}, \qquad \forall m,k : m' \neq m, k' \neq k; \quad \forall i,c,t$$

and

$$p_{m,i,c,t} \leq P_{PF}^{total} \sum_{k=1}^{k=K} z_{m,i,k,c,t}, \qquad \forall m,i,c,t$$

We also found that the following constraints are sufficient to satisfy $C3$ and $C4$ in **Problem O**:

- $$z_{m,i,k,c,t} \geq \frac{A_m - \Omega}{\frac{g_{i,k,c,t}p_{m,i,c,t}}{\sum_{m=1}^{M}\sum_{\forall i' \in \mathcal{S}, i' \neq i} g_{i',k,c,t}p_{m,i',c,t} + \sigma^2} - \Omega}, \quad \forall m,i,k,c,t$$

- $$g_{i,k,c,t} z_{m,i,k,c,t} \leq B_m \left(\sum_{m=1}^{M} \sum_{\substack{\forall i' \in \mathcal{S} \\ i' \neq i}} g_{i',k,c,t} P_{m,i',c,t} + \sigma^2 \right), \qquad \forall m,i,k,c,t$$

- $$\sum_{c=1}^{C}\sum_{t=1}^{T} z_{m,i,k,c,t} \geq z_{m,i,k,c,t} y_m^{min}, \qquad \forall m,i,k,c,t \qquad \forall m,i,k,c,t$$

- $$\sum_{c=1}^{C}\sum_{t=1}^{T} z_{m,i,k,c,t} \leq y_m^{max}, \qquad \forall m,i,k$$

where $A_m = max\left(\left(2^{\frac{F\times R_m}{y_m^{min}\Delta B\Delta T}}-1\right), \gamma^{th}\right)$, $B_m = 2^{\frac{F\times C_m}{y_m^{max}\Delta B\Delta T}}-1$, Ω is a constant strictly greater than A_m, y_m^{min} and y_m^{max} are predetermined integer constants defined as the minimum and maximum number of frequency-time slots respectively that can be allocated to a multicast session.

Finally our optimization problem can be expressed as:

Problem OP1 :

$$\max_{z_{m,i,k,c,t}, p_{m,i,c,t}} \sum_{m=1}^{M}\sum_{i=1}^{S}\sum_{k=1}^{K}\sum_{c=1}^{C}\sum_{t=1}^{T} z_{m,i,k,c,t} \quad \text{(OP1)}$$

s.t.

$$D1 : z_{m,i,k,c,t} \leq \lambda_{m,i,k}, \qquad \forall m, i, k, c, t$$

$$D2 : z_{m,i,k,c,t} + z_{m,i,k',c',t'} \leq 1 + z_{m,i,k,c',t'}, \forall m, i;$$
$$\forall k, k' : k \neq k'; \forall c, c' : c \neq c'; \forall t, t' : t \neq t'$$

$$D3 : z_{m,i,k,c,t} \geq \frac{A_m - \Omega}{\frac{g_{i,k,c,t}p_{m,i,c,t}}{\sum_{m=1}^{M}\sum_{\forall i' \in \mathcal{S}, i' \neq i} g_{i',k,c,t}p_{m,i',c,t} + \sigma^2} - \Omega}, \forall m, i, k, c, t$$

$$D4 : z_{m,i,k,c,t} \leq \frac{B_m\left(\sum_{m=1}^{M}\sum_{\forall i' \in \mathcal{S}, i' \neq i} g_{i',k,c,t}p_{m,i',c,t} + \sigma^2\right)}{g_{i,k,c,t}p_{m,i,c,t}},$$
$$\forall m, i, k, c, t$$

$$D5 : z_{m',i,k',c,t} \leq 1 - z_{m,i,k,c,t}, \ \forall m, k : m' \neq m, k' \neq k; \ \forall i, c, t$$

$$D6 : \sum_{c=1}^{C}\sum_{t=1}^{T} z_{m,i,k,c,t} \geq z_{m,i,k,c,t}y_m^{min}, \qquad \forall m, i, k, c, t$$

$$D7 : \sum_{c=1}^{C}\sum_{t=1}^{T} z_{m,i,k,c,t} \leq y_m^{max}, \qquad \forall m, i, k$$

$$D8 : z_{m,i,k,c,t} \in \{0, 1\}, \qquad \forall m, i, k, c, t$$

$$D9 : \sum_{m=1}^{M}\sum_{i=1}^{S}\sum_{c=1}^{C} p_{m,i,c,t} \leq P_{PF}^{total}, \qquad \forall t$$

$$D10 : p_{m,i,c,t} \leq P_{PF}^{total}\sum_{k=1}^{k=K} z_{m,i,k,c,t}, \qquad \forall m, i, c, t$$

$$D11 : p_{m,i,c,t} \geq 0, \qquad \forall m, i, c, t.$$

In the next section, we explain the proposed solution techniques that we use for **Problem OP1**.

V. Proposed Solution Techniques

To find acceptable bounds in the bounding subroutine of the branch and bound algorithm that solves **Problem OP1**, we construct the Lagrangian relaxation problem **Problem LR**. The relaxed constraints are replaced with a penalty term in the objective function involving the amount of violation of the constraints and their dual variables. For the minimization part, the subgradient algorithm is used to solve for the dual variable vector **u** in **Problem LR** [12], [13].

Problem LR :

$$f(\mathbf{u}, \mathbf{z}, \mathbf{p}) = \min_{\mathbf{u}}\left\{\max_{\mathbf{z},\mathbf{p}}\left(\mathbf{e}.\mathbf{z} + \mathbf{u}\left(\mathbf{b}_1(\mathbf{z}, \mathbf{p}) - \mathbf{b}_2(\mathbf{z}, \mathbf{p})\right)\right)\right\} \quad \text{(LR)}$$

s.t.

$$C^*1 : \mathbf{u} \geq \mathbf{0},$$

$$C^*2 : \mathbf{z}, \mathbf{p} \in F.S.,$$

$$C^*3 : z_{m,i,k,c,t} \in \{0, 1\}, \qquad \forall m, i, k, c, t$$

$$C^*4 : \mathbf{p} \geq \mathbf{0}.$$

where **u** is the dual variable vector, **z** is a vector whose elements are $z_{m,i,k,c,t}$, **p** is a vector whose elements are $p_{m,i,c,t}$, $F.S.$ is the feasible region of the relaxed problem, **e** is a vector of ones, $\mathbf{b}_1(\mathbf{z}, \mathbf{p})$ and $\mathbf{b}_2(\mathbf{z}, \mathbf{p})$ are the right hand side (R.H.S) and left hand side (L.H.S) vectors respectively of the dualized constraints.

For the maximization part, we solve two separate subproblems for the variable vectors **z** and **p**. We use three different configurations of dualized constraints to solve **z**. Each configuration exploits the structure of the relaxed problem depending on the relaxed constraint sets. The basis of the approach is to solve the problem in two phases iteratively where in phase 1, we solve subproblem 1 for **z** while using a constant **p** obtained from the previous iteration of the Lagrangian relaxation problem. In phase 2, the opposite is done by solving subproblem 2 for **p** while fixing **z** at the value obtained from phase1. For both phases, we have linear optimization problems with special structures that we could exploit to solve easily. The subproblem in **z** (phase 1) is the joint scheduling-subchannel allocation Lagrangian relaxation subproblem for each user terminal while the subproblem in **p** (phase 2) is the Lagrangian relaxation subproblem **Problem LR** for the power allocation for each multicast session across the entire HAP service area.

Solving iteratively on two phases can enhance the solution of the maximization part in the Lagrangian problem **Problem LR** or at least does not worsen it. The larger the value of the objective function, for a given dual vector **u**, the smaller the duality gap which means a better bound goodness. For a given dual vector $\mathbf{u}^l$ in a subgradient algorithm iteration l, solving for **z** in phase 1 of the first iteration requires fixing **p** to an arbitrary value $\mathbf{p}^*$ that satisfies the constraints C^*2 and C^*4 of **Problem LR**. The obtained solution for $\mathbf{z}^*$ can then be substituted in **Problem LR** to find a value for **p** which is either a better solution or in the worst case a solution that yields the same value for the objective function obtained from the previous phase. We believe that the new solution for **p** will not be worse than the older one since both solutions are feasible to the Lagrangian relaxation problem **Problem LR**, hence either the new solution is better than the older one or at least is equal to it.

The algorithm is used to continue solving phase 1 and phase 2 iteratively until one of the stopping criteria is satisfied:

1) The duality gap given by:

$$duality\ gap(\%) = 100 - 100\left(\frac{f^*_{i,j}(\mathbf{u}^l, \mathbf{z}, \mathbf{p})}{f_{i,j}(\mathbf{u}^l, \mathbf{z}, \mathbf{p})}\right) \leq \epsilon_o \quad (4)$$

where $f_{i,j}(\mathbf{u}^l, \mathbf{z}, \mathbf{p})$ is the value of the Lagrangian relaxation objective function of **Problem LR** after solving phase i in iteration j of the proposed technique, $f^*_{i,j}(\mathbf{u}^l, \mathbf{z}, \mathbf{p})$ is the best feasible solution obtained and ϵ_o is the acceptable duality gap in percent.
2) The solution does not improve in a specified number of iterations N.
3) The maximum iteration limit, J, is reached.

If there is no improvement achieved after a number of iterations N or if the improvement is slow such that the number of iterations reaches the maximum allowable iterations J before an acceptable duality gap is achieved, the subgradient algorithm can move on to the next iteration $l+1$ to calculate the value of the new dual variable vector $\mathbf{u}^{l+1}$ in its path to find the optimal solution $\mathbf{u}^{**}$ of the dual problem. In the next section we show the different dualization configurations and the corresponding solution methods for subproblem 1 that solves the Lagrangian relaxation problem **Problem LR** for $\mathbf{z}$ (phase1).

A. *Proposed Solution Methods for Phase 1 and Their Corresponding Dualizations*

The following are the three different proposed dualization configurations and their corresponding proposed solution methods for the phase 1 subproblem to solve for $\mathbf{z}$

- **Method1**: We dualize all the constraint sets except constraint $D7$,

$$\sum_{c=1}^{C}\sum_{t=1}^{T} z_{m,i,k,c,t} \leq y_m^{max}, \qquad \forall m, i, k$$

and $D8$, $z_{m,i,k,c,t} \in \{0, 1\}$, $\forall m, i, k, c, t$. The Lagrangian relaxation problem can then be further broken into $M \times S \times K$ independent sub-problems that can all be solved separately. Those subproblems can each be solved to optimality using a heap sort algorithm to find the highest y_m^{max} objective function coefficients of the Lagrangian relaxation problem. The strategy is to go through the coefficients once, and keep a list of the highest y_m^{max} elements found so far. This is done efficiently by always knowing the smallest element in this top-y_m^{max}, so that it can be possibly replaced by a larger one. The heap structure makes it easy to maintain this list without wasting any effort. This way does enough of the sort to find the smallest element, and that is why it is fast. The worst case complexity is $O\left(MSKCT \log\left(\max_m(y_m^{max})\right)\right)$ [21].

- **Method2**: Dualizes all constraints except $D1$ and either $D2$ or $D3$ in problem **Problem LR** or both. The Lagrangian problem can then be solved by setting to '1' those $z_{m,i,k,c,t}$ which satisfy:
  i. $\lambda_{m,i,k}$ and
  ii. $g_{i,k,c,t}p_{m,i,c,t} \leq B_m(\sum_{m=1}^{M}\sum_{\forall i' \in \mathcal{S}, i' \neq i} g_{i',k,c,t}p_{m,i',c,t} + \sigma^2)$
  and/or
  iii. $g_{i,k,c,t}p_{m,i,c,t} \geq A_m(\sum_{m=1}^{M}\sum_{\forall i' \in \mathcal{S}, i' \neq i} g_{i',k,c,t}p_{m,i',c,t} + \sigma^2)$.

  This can be done for all $M \times S \times K \times C \times T$ $z_{m,i,k,c,t}$ variables separately and the solution to the Lagrangian relaxation problem will still be optimal. The worst case complexity is derived to be $O(MS^2KCT)$.
- **Method3**: Relaxes only one constraint set of those which have $p_{m,i,c,t}$, $D2$ and/or $D3$ in problem **Problem LR**, to bring in $p_{m,i,c,t}$ in the objective function so that in phase 2, the sub-problem becomes a simple LP with a linear objective function in $p_{m,i,c,t}$. This problem can be solved using branch and bound with LP relaxation to obtain the bounds at each node in the tree [22]. The BnB algorithm may stop either after the optimal solution to the Lagrangian relaxation problem is obtained or at an acceptably close to optimal feasible solution. The worst case complexity of BnB occurs when it requires examination of all the nodes in its tree. In this case its complexity is $O(2^{MSKCT})$.

VI. Power Allocation Subproblem (Phase 2) and Proposed Solution

Solving for the power vector $\mathbf{p}$ could be done using the simplex algorithm [22] for any of the three proposed dual configurations. This is because for the fixed value of the vector $\mathbf{z}$ obtained from phase 1, the Lagrangian relaxation **Problem LR** is a linear program in $\mathbf{p}$. However for dualization configuration 1 (method 1), a greedy algorithm can be used to solve the special structured linear program which is described in this section. This has a much lower worst case complexity as we will see compared to the exponential worst case complexity of simplex algorithm.

For a subgradient algorithm iteration l, for a given dual variable vector $\mathbf{u}^l$, a power allocation linear optimization problem is given by problem:

Problem PW :

$$\max\ \mathbf{e}\mathbf{z}^* + \mathbf{u}^l\mathbf{b}\left(z^*_{m,i,k,c,t}\right) - \mathbf{u}^l_1\widehat{\mathbf{A}}_1\mathbf{z}^* - \mathbf{u}^l_2\widehat{\mathbf{A}}^p_2\left(z^*_{m,i,k,c,t}\right)\mathbf{p} \quad \text{(PW)}$$

s.t.

$$L1^* : \sum_{m=1}^{M}\sum_{i=1}^{S}\sum_{c=1}^{C} p_{m,i,c,t} \leq P_{PF}^{total}, \qquad \forall t$$

$$L2^* : p_{m,i,c,t} \geq 0, \qquad \forall m, i, c, t$$

where

$$\mathbf{b}(z_{m,i,k,c,t}) = \begin{bmatrix} \widehat{\mathbf{b}_1} \\ \widehat{\mathbf{b}_2}(z_{m,i,k,c,t}) \end{bmatrix},$$

$$\widehat{\mathbf{b}_1} = \begin{bmatrix} \tilde{\mathbf{b}}_1 \\ \tilde{\mathbf{b}}_2 \\ \tilde{\mathbf{b}}_5 \\ \tilde{\mathbf{b}}_6 \end{bmatrix}, \widehat{\mathbf{b}_2}(z_{m,i,k,c,t}) = \begin{bmatrix} \tilde{\mathbf{b}}_3 \\ \tilde{\mathbf{b}}_4 \\ \tilde{\mathbf{b}}_{10}(z_{m,i,k,c,t}), \end{bmatrix}$$

$$\widehat{\mathbf{A}}_1 = \begin{bmatrix} \tilde{\mathbf{A}}_1 \\ \tilde{\mathbf{A}}_2 \\ \tilde{\mathbf{A}}_5 \\ \tilde{\mathbf{A}}_6 \end{bmatrix}, \widehat{\mathbf{A}}_2(z_{m,i,k,c,t}) = \begin{bmatrix} \tilde{\mathbf{A}}_3(z_{m,i,k,c,t}) \\ \tilde{\mathbf{A}}_4(z_{m,i,k,c,t}) \\ \tilde{\mathbf{A}}_{10} \end{bmatrix}, \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}.$$

$\mathbf{e}$ is a vector of ones, $\tilde{\mathbf{b}}_1, \tilde{\mathbf{b}}_2, \tilde{\mathbf{b}}_5, \tilde{\mathbf{b}}_6$,the right hand side (R.H.S) vectors of constraints $D1, D2, D5, D6$ in the original optimization problem **Problem OP1** and $\tilde{\mathbf{A}}_1, \tilde{\mathbf{A}}_2, \tilde{\mathbf{A}}_5, \tilde{\mathbf{A}}_6$ are their corresponding left hand side (L.H.S) constraint coefficient matrices. The vectors $\tilde{\mathbf{b}}_3^p, \tilde{\mathbf{b}}_4^p, \tilde{\mathbf{b}}_{10}^p(z_{m,i,k,c,t})$ are R.H.S vectors of the constraints $D3, D4, D10$ when written in the form:

$$D3^p : -g_{i,k,c,t}p_{m,i,c,t}z_{m,i,k,c,t} + A_m \left(\sum_{m=1}^{M} \sum_{\substack{\forall i' \in \mathcal{S} \\ i' \neq i}} g_{i',k,c,t}p_{m,i',c,t} + \sigma^2 \right) \leq 0, \qquad \forall m, i, k, c, t \tag{5}$$

$$D4^p : g_{i,k,c,t}z_{m,i,k,c,t}p_{m,i,c,t} - B_m \left(\left(\sum_{m=1}^{M} \sum_{\substack{\forall i' \in \mathcal{S} \\ i' \neq i}} g_{i',k,c,t}p_{m,i',c,t} + \sigma^2 \right) \right) \leq 0, \quad \forall m, i, k, c, t \tag{6}$$

$$D10^p : p_{m,i,c,t} \leq P_{PF}^{total} \sum_{k=1}^{k=K} z_{m,i,k,c,t}, \qquad \forall m, i, c, t \tag{7}$$

and $\tilde{\mathbf{A}}_3^p(z_{m,i,k,c,t}), \tilde{\mathbf{A}}_4^p(z_{m,i,k,c,t}), \tilde{\mathbf{A}}_{10}^p$ are their L.H.S coefficient matrices respectively.

Swapping columns of the functional constraint matrix of **Problem PW** and its objective function terms as expressed by the indices of $p_{m,i,c,t}$ to $p_{t,m,i,c}$ gives us a T independent and separate linear programs that can be solved in parallel. Furthermore, by inspection for each of the t subproblems, if the L.H.S and R.H.S of each of the functional constraint are divided by the R.H.S of the constraint, which is P_{PF}^{total}, the functional constraint for subproblem t can be rewritten as $\sum_{m=1}^{M} \sum_{i=1}^{S} \sum_{c=1}^{C} \tilde{p}_{m,i,c,t} \leq 1, \ \forall t$ where $\tilde{p}_{m,i,c,t} = \frac{p_{m,i,c,t}}{P_{PF}^{total}}$. The decision variable vector can also be changed to $\tilde{\mathbf{p}}$ and an equivalent objective function is therefore

$$\max \left(\mathbf{e}\mathbf{z}^* + \mathbf{u}^l \mathbf{b} \left(z_{m,i,k,c,t}^* \right) - \mathbf{u}_1^l \widehat{\mathbf{A}}_1 \mathbf{z}^* - P_{PF}^{total} \mathbf{u}_2^l \widehat{\mathbf{A}}_2^p \left(z_{m,i,k,c,t}^* \right) \tilde{\mathbf{p}} \right).$$

Since each of the T subproblems the L.H.S coefficients of the functional constraints are all ones, then the elements of the objective function coefficient vector (call it ρ) can be reordered. This can be implemented by the the **Merge − Sort** algorithm described earlier. The solution procedure for each subproblem t in **Problem PW** can be implemented as follows:

$\overline{y}_m^{max} \leftarrow y_m^{max}$
$\tilde{\mathbf{p}} \leftarrow \mathbf{0}$
$\mathbf{Merge - Sort}(\rho)$
for $(j = 1; j \leq N; j++)$
  if $\left(\overline{y}_m^{max} \geq 1\right)$
    $\tilde{\mathbf{p}}_{m,i}(j) \leftarrow 1$
  else
    $\tilde{\mathbf{p}}_{m,i}(j) \leftarrow \overline{y}_m^{max}$
  end if
  $\overline{y}_m^{max} \leftarrow \overline{y}_m^{max} - \tilde{\mathbf{p}}_{m,i}(j)$
  if $\left(\overline{y}_m^{max} \leq 0\right)$
    exit
  end if
end for.

The worst case complexity analysis for this algorithm is simple, knowing that the **Merge − Sort** has a worst case time complexity of $O(Nlog(N))$, where $N = CT$. Besides that, the simple 'for loop' shows and the initial vector assignment both have a linear time complexity $O(N)$. It was easily shown that the column swapping algorithm has a worst case complexity $O(MSN)$. Therefore, the overall complexity of the algorithm is hence $O(Nlog(N) + (MSN) + N)$ for the entire solution procedure for **Problem PW**. This is much lower than the exponential worst case time complexity of the simplex algorithm which happen if the simplex algorithm has to test all the corner points of the feasible region polyhedron [22].

VII Solving the Dual Problem

To solve for the dual variable vector $\mathbf{u}$ in **Problem LR**, t h e subgradient algorithm is used to calculate the search direction and step size ω_l at each iteration l. As in [12] and [13], the direction is given by the vector $(\mathbf{b}_1(\mathbf{z}, \mathbf{p}) - \mathbf{b}_2(\mathbf{z}, \mathbf{p}))$ obtained from the relaxed constraints. The formula used to determine the sequence of values of the dual variable vector $\mathbf{u}^l$ in every iteration is given by [13]:

$$\mathbf{u}^{l+1} = max \left\{ 0, \mathbf{u}^l - \omega_l \left(\mathbf{b}_1(\mathbf{z}^*, \mathbf{p}^*) - \mathbf{b}_2(\mathbf{z}^*, \mathbf{p}^*) \right) \right\}. \tag{8}$$

For the step size ω_l, we compare three different step size rules, these are:

1) **Step size rule 1**: Square summable but not summable [14] which satisfies:

$$\sum_{l=1}^{\infty} \omega_l^2 < \infty \ and \ \sum_{l=1}^{\infty} \omega_l = \infty$$

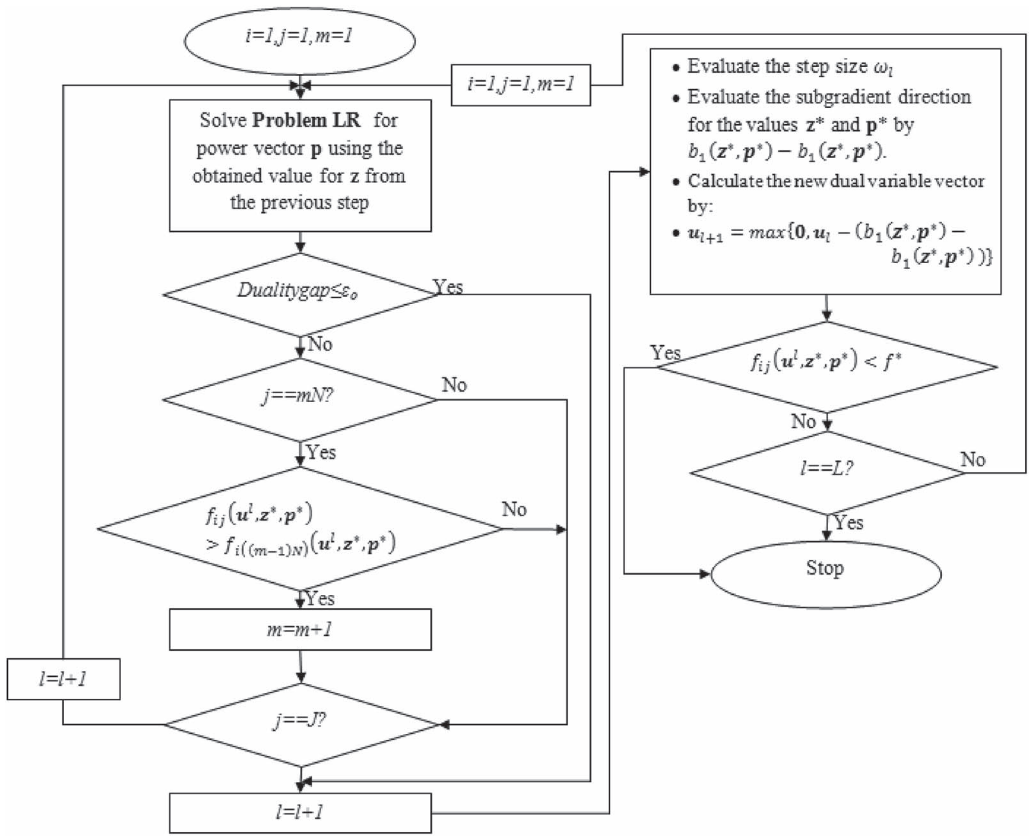


Fig. 1. Bounding subroutine flowchart.

where ω_l is the step size at the l^{th} iteration. Therefore we choose

$$\omega_l = \frac{1}{l}. \tag{9}$$

2) **Step size rule 2**: Nonsummable diminishing [14] which satisfies:

$$\lim_{l\to\infty} \omega_l = 0 \ and \ \sum_{k=1}^{\infty} \omega_l = \infty$$

and hence we choose

$$\omega_l = \frac{1}{\sqrt{l}}. \tag{10}$$

3) **Step size rule 3**: The formula that Fisher suggested in his papers [12] and [13]:

$$\omega_l = \frac{\rho_l \left(f(\boldsymbol{u}^l) - f^*\right)}{\|\mathbf{b}_1(\mathbf{z}, \mathbf{p}) - \mathbf{b}_2(\mathbf{z}, \mathbf{p})\|^2} \tag{11}$$

where ρ_l is a scalar satisfying $0 < \rho_l \leq 2$.

The termination conditions for the sub-gradient algorithm are chosen to be any of the following that first takes place:

1) $f(\boldsymbol{u}) < f^*$ where f^* is the best feasible solution found.
2) An acceptable duality gap (%).
3) A maximum iteration limit L.

The entire procedure used to find a suitable bound for the problem **Problem OP1** is given in Fig. 1.

VIII. NUumerical and Simulation Results

In this section we illustrate some numerical and simulation results for both subproblems 1 and 2. First we perform some numerical experiments with parameters that yield a medium sized optimization problem. The aim of these experiments is to numerically compare the bounds obtained by different solution methods, find the most suitable step size formula to use and to find out whether the initial values of the dual variable vectors affect the bounds. Based on the obtained results, we use the step size formula that performed best in a realistic simulation scenario that is described in detail later to numerically observe and compare the results of the proposed Lagranigan relaxation based solution methods in terms of bounds.

For initial numerical experiments for subproblem1, the bound goodness is calculated as

$$Bound\_Goodness = 100 \times \left(\frac{f^*}{f(\boldsymbol{u})}\right), \tag{12}$$

where

- $f(\boldsymbol{u})$ is the objective function value of the dual problem of phase 1 with $\mathbf{p}$ fixed to a value $\mathbf{p}^*$,
- f^* is the value of the objective function of the primal problem **Problem OP1** at its optimal feasible solution for a fixed power vector value $\mathbf{p}^*$.

Table II
EXPERIMENTAL VALUES FOR MODEL PARAMETERS

Model Parameters	Values
Noise power spectral density (N_o)	-173 dBm/Hz
OFDMA frame length (F)	10 ms
Minimum capacity requirements (R_m)	R_1=0.5 Mbps, R_2=0.75 Mbps
Maximum capacity requirements (C_m)	C_1=6 Mbps, C_2=7 Mbps
Total Bandwidth	1MHz
Total HAP Power	120 Watt
Target BER	10^{-6}
Constellation Size	2
Carrier Frequency	28 GHz
$\lambda_{m,i,k}$	$\lambda_{1,1,1}$= $\lambda_{1,1,2}$=$\lambda_{2,2,3}$= $\lambda_{2,2,4}$=1 and zero for others

The optimization problem for $M = 2$, $S = 2$, $K = 4$, $C = 2$, $T = 2$, $y_m^{min} = 1\ \forall\, m$, $y_m^{max} = 3\ \forall\, m$ is considered and the rest of the numerical parameters of the problem are shown in Table II. The parameter ρ_l of **step size rule 3** begins with a value of 2 and is reduced by half if the bound does not decrease for 500 iterations. LP relaxation is used as a reference scheme, where the integer constraint of $z_{m,i,k,c,t}$ is relaxed to take any values in the range $0 \leq z_{m,i,k,c,t} \leq 1$.

Fig. 2(a) shows the goodness of the bounds obtained using Methods 1, 2, and 3 by using **step size rule 1** compared to that of LP relaxation when the initial dual variable vector $\boldsymbol{u}$ is set to $\mathbf{0}$. For all the three methods, the subgradient algorithm converged to bounds with the goodness shown on the chart and remained at those values for thousands of iterations before the algorithm reached its maximum iteration limit. They all seem to perform almost at least as the LP relaxation in the worst case. Fig. 2(b) and (c) are similar charts for **step size rules 2 and 3**, respectively. We can see that the bounds provided by **step size rule 3** are the worst. This is because there are no guarantees that they converge to the optimal solution $f(\boldsymbol{u}^*)$ since according to [12], **step size rule 3** does not satisfy the sufficient conditions of correct convergence given by $\lim_{l\to\infty} \omega_l = 0$ and $\sum_{l=1}^{\infty} \omega_l = \infty$. As a result, some of the relaxations perform much poorer than LP relaxation using this step size rule as the chart in Fig. 2(c) shows. The charts show that among the three proposed solution methods, method 1 has the lowest bound goodness and, depending upon the constraint sets relaxed, the highest is either Method 2 or Method 3. However, it is worth mentioning that according to the worst case complexity mentioned previously in this paper, Method 1 has the lowest complexity. Therefore, there is a trade-off between Methods 1 and 2 in terms of bound goodness and worst case complexity.

Fig. 3(a)–(c) show the bound goodness for each of the proposed solution methods versus the initial values of the dual variables for step size rules 1, 2 and 3 respectively. For Fig. 3(a), the difference in bound goodness for different initial values of **u** is relatively large because the convergence using step **size rule 1** is not as fast as that in Fig. 3(b) (**step size rule 2**). The faster convergence for step size rule 2 compared to that of 1 is expected due to the higher dampness. Fig. 4(a) and (b) support our argument which give the absolute value of

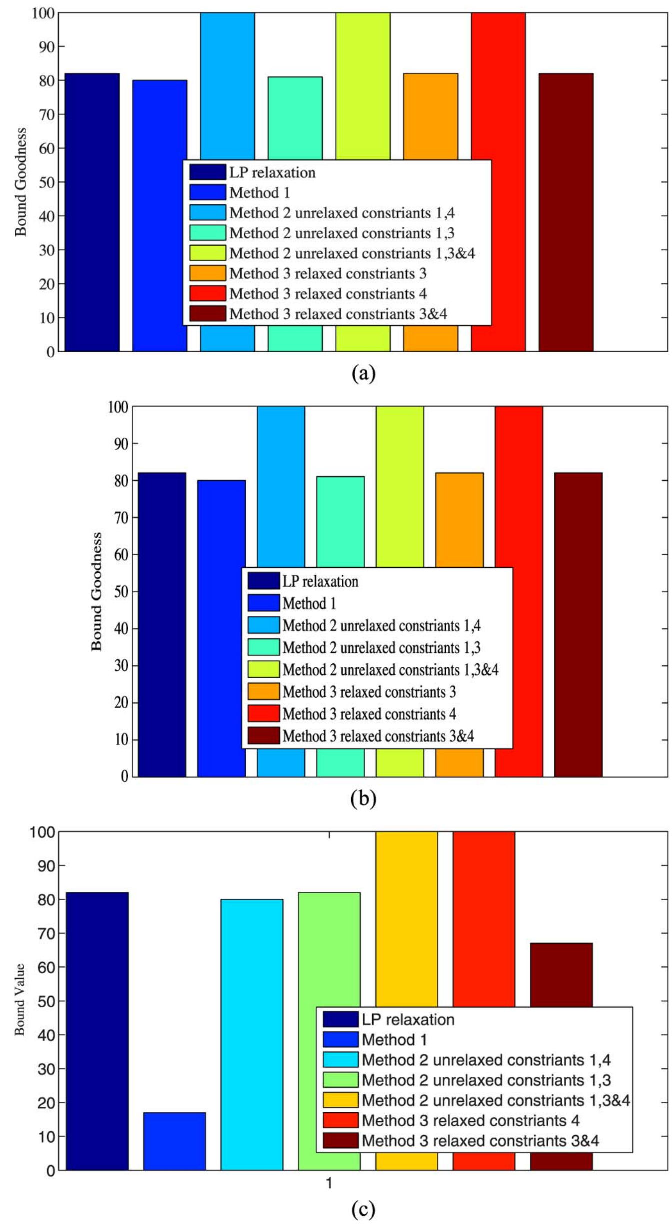
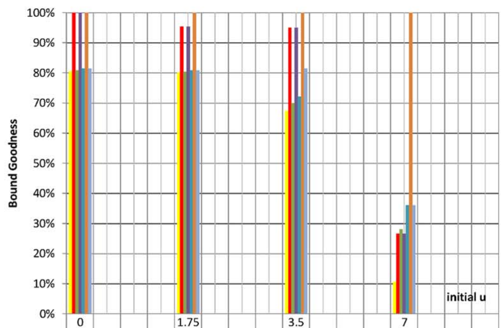


Fig. 2. Goodness of bounds for the three proposed Lagrangian relaxation based solution methods for phase 1. (a) Step size rule no. 1. (b) Step size rule no. 2. (c) Step size rule no. 3.

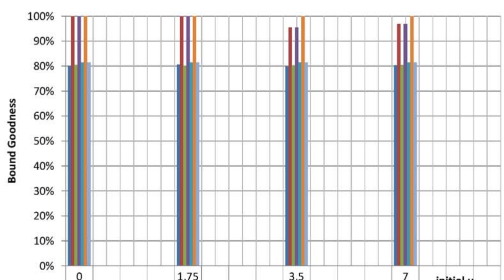
the obtained bounds versus the number of iterations of the subgradient algorithm for step size rules 1 and 2 respectively. For those initial points that do not lead to convergence in the iteration limit mentioned previously, convergence is expected for a larger iteration limit.

For **step size rule 3** in Fig. 3(c), the difference in the goodness of bounds due to different initial dual variable values is huge. The reason is that as Fig. 4(c) shows, convergence is too slow to the extent that leads us to question if it ever converges to the same solution. We believe that this is due to the fact that it does not satisfy the sufficient conditions for convergence as Fisher mentioned in his paper [12]. **Step size rule 2** seems to converge quickly to the same solution even if the initial point differs and hence is the best choice for our problem.



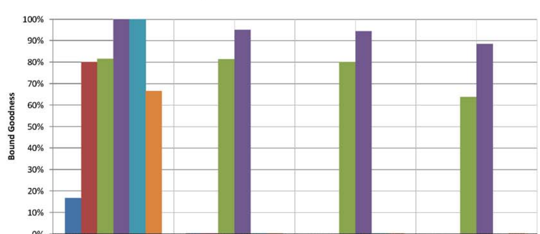
initial u	0	1.75	3.5	7
Method1	80%	80%	67%	11%
Method 2 Unrelaxed Constraint Sets 1,4	100%	95%	95%	27%
Method 2 , Unrelaxed Constraints 1,3	81%	80%	70%	28%
Method2 Unrelaxed Constraints (1,3,4)	100%	95%	95%	27%
Method 3, Relaxed Constraint set 3	82%	81%	72%	36%
Method 3, Relaxed Constraints 4	100%	100%	100%	100%
Method 3, relaxed constraints 3&4	82%	81%	82%	36%

(a)



initial u	0	1.75	3.5	7
Method 1	80%	81%	80%	80%
Method 2, Unrelaxed constraints 1,4	100%	100%	96%	97%
Method 2 , Unrelaxed Constraints 1,3	81%	80%	80%	81%
method 2 , Unrelaxed Constraints 1,3 and 4	100%	100%	96%	97%
Method 3 ,relaxed cosntraint set 3	82%	81%	81%	81%
Method 3, relaxed constraints 4	100%	100%	100%	100%
Method 3, Relaxed constraints 3 & 4	82%	81%	81%	81%

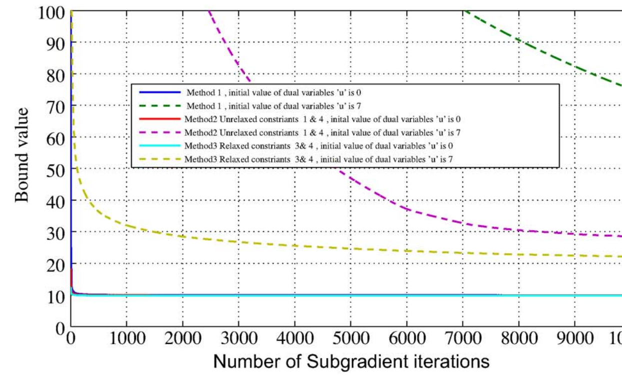
(b)



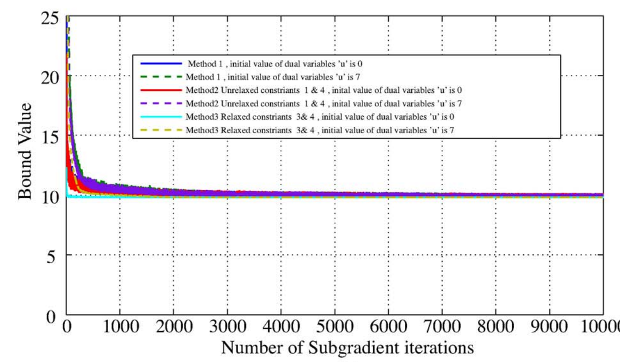
Initial u	0	1.75	3.5	7
Method 1	17%	0%	0%	0%
Method 2, Unrelaxed constraints 1&4	80%	0%	0%	0%
Method 2 Unrelaxed constraints 1&3	82%	81%	80%	64%
Method 2, Unrelaxed constraints 1,3 &4	100%	95%	94%	89%
Method 3,Relaxed Cosntriant set 4	100%	0%	0%	0%
Method 3, Relaxed Constriant sets 3 & 4	67%	0%	0%	0%

(c)

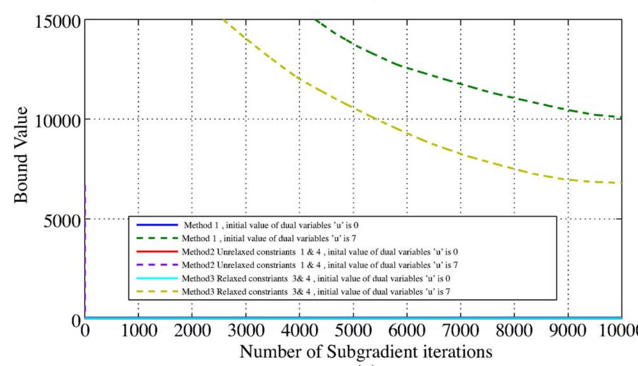
Fig. 3. Goodness of bounds for the three proposed Lagrangian relaxation based solution methods at different initial dual variable values. (a) Step size rule no.1. (b) Step size rule no. 2. (c) Step size rule no. 3.



(a)



(b)



(c)

Fig. 4. Bounds vs subgradient algorithm iterations for the three proposed Lagrangian relaxation based solution method. (a) Step size rule no.1. (b) Step size rule no. 2. (c) Step size rule no. 3.

We then vary the optimization problem size by varying the number of users terminals (K) in the range 3–7. All the other model parameters that we used remain the same as those in Table II. Consider **step size rules 1 and 2**, Fig. 5(a) and (b) show that proposed Method 2 performs better than others in terms of bound goodness. For **step size rule 3**, the bound goodness is poor compared to **step size rules 1 and 2** as Fig. 5(c) shows. These results are consistent with the previous results.

A more realistic simulation was conducted for a cell surrounded by its six first tier neighbor cells of the HAP service area (i.e. 7 cells total). Each cell has a radius of 5 km and the UTs are uniformly distributed around its center. The HAP subplatform (SS) point coincides with the center of the middle cell. The following parameters and values are used in our simulation:

- HAP height is 20 Km.
- Total platform power $P_{PF}^{Total} = 7W$.
- Initial power distribution $p_{m,i,c,t} = P_{PF}^{Total}/M.S.C$ is uniform across all subchannels, cells and multicast sessions.
- Four subchannels.
- Two time slots.
- Two multicast sessions in the service area.
- $R_m = 256$ Kbps and $C_m = 10$ Mbps.
- $y_m^{min} = 1$ for $m = 1, 2$, while $y_1^{max} = 6$ and $y_2^{max} = 4$.
- The values of noise power spectral density, OFDMA frame length, total bandwidth, target BER and carrier frequency are in Table II.

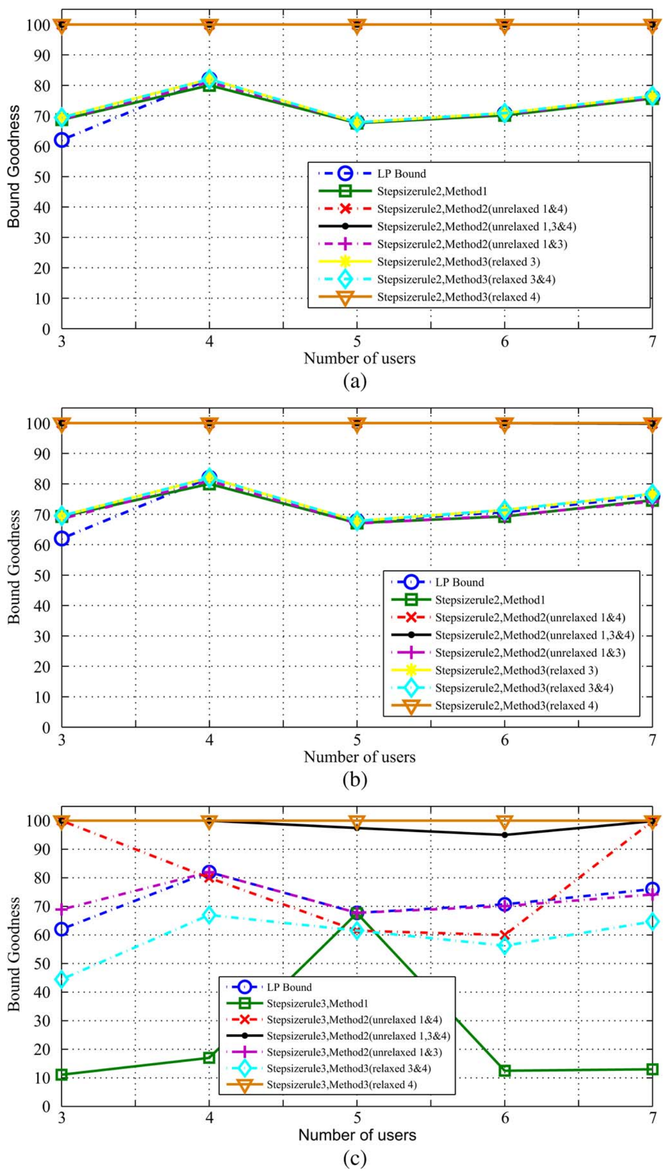


Fig. 5. Bound goodness for the three proposed methods versus the number of users (K). (a) Step size rule no.1. (b) Step size rule no. 2. (c) Step size rule no. 3.

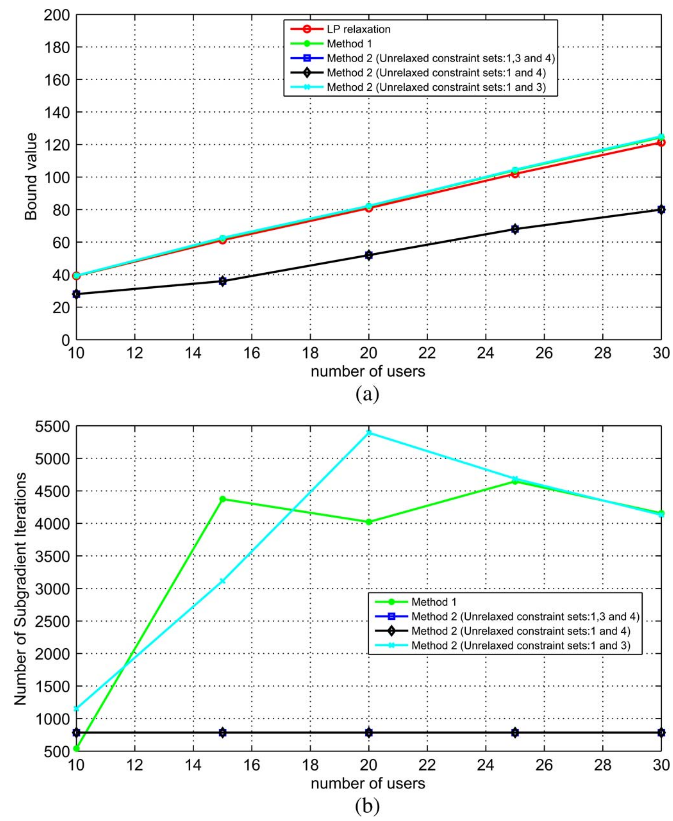


Fig. 6. Bound values and the number of required subgradient iterations to obtain the bounds (a) bound value (b) number of subgradient iterations.

For the channel model, we considered the following:

- Free space path loss was considered for the large scale fading, (i.e. path-loss exponent of 2).
- a rain attenuation model over the first 3 km of height as the one used in [18] is used:

$$A(d_k) = 10^{\left(\frac{3d_k\chi}{10h}\right)} \tag{13}$$

  where h is the HAP height, χ is the attenuation through the clouds and rain in dB/km.
- For the small scale fading we use Ricean fading with a Rice factor of 20 dB.

For the HAP antennas we use the aperture antenna model that was used in [20] in which the gain of the main lobe with respect to the boresight angle is defined as:

$$G_H(\Psi) = Ap_{eff} \cdot cos(\Psi)^n \frac{32log2}{2\left(2arccos\left(\sqrt[n]{0.5}\right)\right)^2} \tag{14}$$

where Ap_{eff} is the aperture efficiency which we consider to be unity and n is the roll-off factor of the antenna. The roll-off factor value is chosen to maximize the gain at the cell borders as in [20]. The side lobes are modeled by a flat level at −40 dB.

Each user is considered to have parabolic antenna, whose gain when receiving a transmission from the HAP antenna of the cell to which the user belongs to is given by:

$$G_u = A^u_{eff}\left(\frac{\pi D f_c}{C_{light}}\right)^2 \tag{15}$$

where A^u_{eff} is the aperture efficiency of the parabolic antenna and we take that to be unity in our simulation and D is the diameter of the parabolic reflector. Any received transmission from any other HAP antenna of the other cells falls in the sidelobes of the user antenna which is modelled by a flat level of 0 dB.

Using the observations from the previous results of smaller problem sizes, we use **step size rule 2** for its quick and correct convergence regardless of the initial dual variable vector values. Fig. 6(a) and (b), show the bound values and the required number of subgradient iterations for dual configuration 1 (Method 1) and dual configuration 2 (Method 2) when the number of users are varied from in the range 10–30 at increments of five. Method 3 was not shown since its is BnB based have a much higher (exponential) complexity compared to the other Method 1 and Method2 and cannot be implemented for real time resource allocation. In Fig. 6(a), a low bound is tighter than a higher one and hence is closer to the solution.

The results for the bounds obtained are consistent with the previous results that showed the bound goodness for each method. Also, the results in Fig. 6(b). The number of subgradient algorithm iterations required to converge to the bound value was lowest for dual configuration 2 (Method 2) when the undualized constraints are $D4$ *and/or* $D3$ *and* $D1$. However in a single iteration, the worst case complexity for dual configuration 1 (Method 1) is lower ($O\left(MSKCT \log\left(\max_{m}\left(y_m^{max}\right)\right)\right)$ compared to $O(MS^2KCT)$).

In a second experiment that we conducted, we show the value of the objective function versus the number of iterations between phases 1 and 2. Since this is a Lagrangian relaxation problem, we set initial values for the dual variable vector $\mathbf{u}$ and try $\mathbf{u} = \mathbf{1}$, $\mathbf{u} = \{2, 2 \ldots 2\}$, $\mathbf{u} = \{3, 3 \ldots 3\}$. The reason for comparing the results for different dual variable values is that these and the initial power configurations are the only parameters whose values can differ for each subgradient iteration. We therefore want to make sure that the behavior is the same for different values of those parameters. For power initialization, we use $p_{m,i,c,t} = \frac{P_{PF}^{total}}{pm \times M \times C \times S}$ where pm is an integer power multiplier that we vary in the range 1–4.

The results provided in Fig. 7 show that for a given dual variable vector $\mathbf{u}$:

1) The value of the Lagrangian relaxation objective function monotonically increases.
2) The solution is bounded since the value if the Lagrangian relaxation objective function stops increasing after a finite number of iterations.
3) No matter what the initial power multiplier value pm is, the the procedure converges to the same solution.
4) The number of required iterations to reach convergence in the experiments that we conducted was no more than three iterations only.

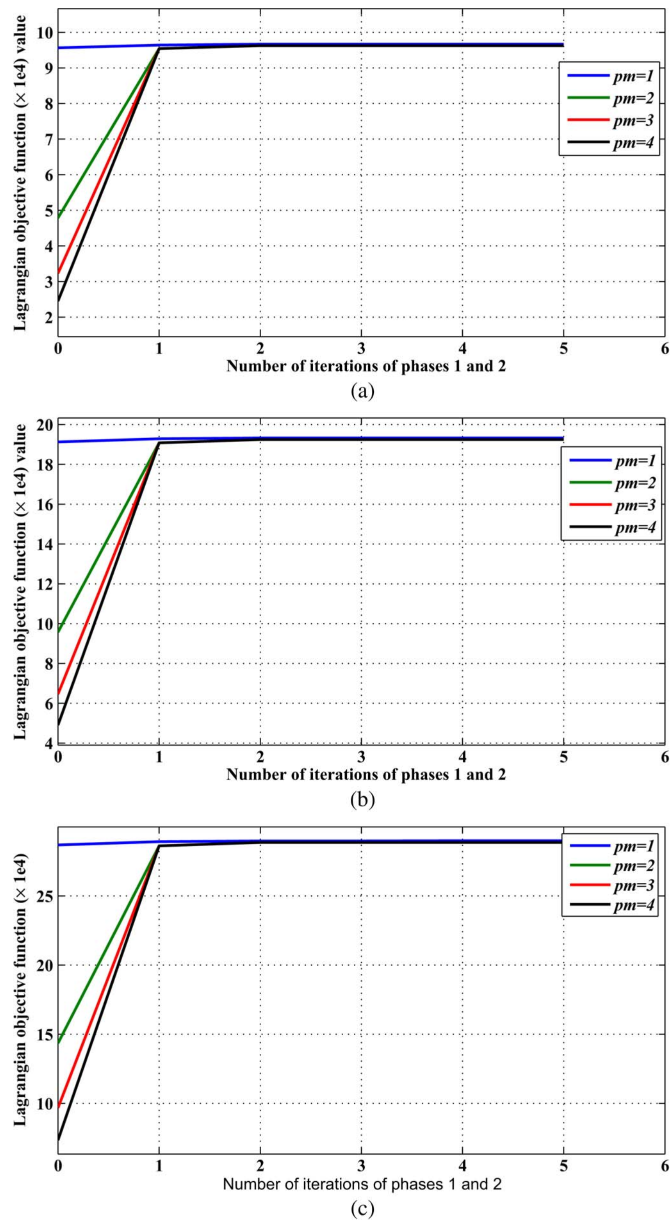


Fig. 7. The objective function value versus iteration no. at different initial dual variable values. (a) $\mathbf{u} = \mathbf{1}$. (b) $\mathbf{u} = \{2, 2, \ldots, 2\}$. (c) $\mathbf{u} = \{3, 3, \ldots, 3\}$.

IX. Conclusion

In this paper, we focused on RRA for OFDMA based HAP system for multicast transmissions in the downlink in order to maximize the number of user terminals being served. We formulated an optimization problem for our considered system that turned out to be an MINLP. We suggested a technique to obtain bounds in BnB to solve the problem based on Lagrangian relaxation. We explained in details different algorithms and methods to solve the $\mathbf{z}$ and $\mathbf{p}$ Lagrangian subproblems. Numerical and simulation results showed the goodnesses of the bounds. Dual configuration 2 (method 2) and dual configuration 3 (method 3) gave the best bounds while dual configuration 1 (method 1) gave acceptable bounds that are almost equal to those given by LP relaxation. Also, we used three different step size rules for the subgradient algorithm which was used to find the dual variable vector $\mathbf{u}$ and numerically compared their bound goodness. **Step size rule 2** gave correct and quicker convergence compared to the other two.

Although dual configuration 2 (Method 2) gave better bounds than method 1, we showed that Method 1 can be implemented with lower complexity per subgradient iteration. Lower complexity is important to us due to the too quick radio link condition variations. Therefore, Method 1 seems a reasonable choice with acceptable bounds. Using, dual configuration 1 (Method1) for phase 1, we were able to show how a greedy algorithm can solve for $\mathbf{p}$ in **Problem LR**, a linear program, with a lower worst case complexity than simplex. We gave some results that showed that both phases 1 and 2 converge to a solution (a local solution at least) in three iterations.

For future work, energy efficiency and delay can be considered in our problem as these were not considered in this paper.

References

[1] A. Mohammed, A. Mehmood, F. Pavlidou, and M. Mohorcic, "The role of high-altitude platforms (HAPs) in the global wireless connectivity," *Proc. IEEE*, vol. 99, no. 11, pp. 1939–1953, Nov. 2011.

[2] G. Araniti, A. Molinaro, and A. Iera, "Multicast in terrestrial-hap systems," in *Proc. IEEE 66th VTC—Fall*, 2007, pp. 154–158.

[3] G. Araniti, A. Iera, and A. Molinaro, "The role of haps in supporting multimedia broadcast and multicast services in terrestrial-satellite integrated systems," *Wireless Pers. Commun.*, vol. 32, no. 3/4, pp. 195– 213, Feb. 2005.

[4] A. Abrardo and D. Sennati, "Centralized radio resource management strategies with heterogeneous traffics in haps wcdma cellular systems," *IEICE Trans. Commun.*, vol. 86, no. 3, pp. 1040–1049, Mar. 2003.

[5] Y. C. Foo, W. L. Lim, and R. Tafazolli, "Centralized downlink call admission control for high altitude platform station umts with onboard power resource sharing," in *Proc. 56th VTC—Fall*, 2002, vol. 1, pp. 549–553.

[6] Y. C. Foo and W. L. Lim, "Speed and direction adaptive call admission control for high altitude platform station (HAPS) UMTS," in *Proc. IEEE MILCOM*, 2005, pp. 2182–2188.

[7] Y. Liu, D. Grace, and P. D. Mitchell, "Exploiting platform diversity for gos improvement for users with different high altitude platform availability," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 196–203, Jan. 2009.

[8] L. Zhao, L. Cong, F. Liu, K. Yang, and H. Zhang, "Joint time-frequency-power resource allocation for low-medium-altitude platforms-based wimax networks," *IET Commun.*, vol. 5, no. 7, pp. 967–974, May 2011.

[9] L. Zhao, C. Zhang, H. Zhang, X. Li, and L. Hanzo, "Power-efficient radio resource allocation for low-medium-altitude aerial platform based TD-LTE networks," in *Proc. IEEE VTC*, 2012, pp. 1–5.

[10] J. Lee and S. Leyffer, *Mixed Integer Nonlinear Programming*. New York, NY, USA: Springer-Verlag, 2012.

[11] L. Hillier, *Introduction to Operations Research*. New York, NY, USA: McGraw-Hill, 2001.

[12] M. L. Fisher, "The lagrangian relaxation method for solving integer programming problems," *Manage. Sci.*, vol. 27, no. 1, pp. 1–18, 1981.

[13] M. L. Fisher, "An applications oriented guide to lagrangian relaxation," *Interfaces*, vol. 15, no. 2, pp. 10–21, Mar./Apr. 1985.

[14] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," Stanford Univ., Stanford, CA, USA, Lecture Notes EE392o, 2003, vol. 2004.

[15] R. O. Afolabi, A. Dadlani, and K. Kim, "Multicast scheduling and resource allocation algorithms for ofdma-based systems: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 240–254, 1st Quart. 2013.

[16] P. K. Gopala and H. El Gamal, "On the throughput-delay tradeoff in cellular multicast," in *Proc. IEEE Int. Conf. Wireless Netw., Commun. Mobile Comput.*, 2005, vol. 2, pp. 1401–1406.

[17] H. Won *et al.*, "Multicast scheduling in cellular data networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4540–4549, Sep. 2009.

[18] D. A. Pearce and D. Grace, "Optimum antenna configurations for millimetre-wave communications from high-altitude platforms," *IET Commun.*, vol. 1, no. 3, pp. 359–364, Jun. 2007.

[19] E. Falletti, M. Laddomada, M. Mondin, and F. Sellone, "Integrated services from high-altitude platforms: A flexible communication system," *IEEE Commun. Mag.*, vol. 44, no. 2, pp. 85–94, Feb. 2006.

[20] J. Thornton, D. Grace, M. H. Capstick, and T. C. Tozer, "Optimizing an array of antennas for cellular coverage from a high altitude platform," *IEEE Trans. Wireless Commun.*, vol. 2, no. 3, pp. 484–492, May 2003.

[21] [Online]. Available: http://stevehanov.ca/blog/index.php?id=122

[22] P. A. Jensen and J. F. Bard, *Operations Research Models and Methods*. Hoboken, NJ, USA: Wiley, 2003.

Ahmed Ibrahim received the B.Sc. and M.Sc. degrees in electronics and telecommunications engineering from the Arab Academy of Science, Technology and Maritime Transport (AASTMT), Cairo, Egypt, in 2006 and 2010, respectively. He is pursuing his Ph.D. degree at the University of Manitoba (UoM) in the Electrical and Computer Engineering Department.

During 2006–2010, he worked at AASTMT as a Graduate Teaching Assistant and as a Research Student towards his M.Sc. degree. From 2010 to 2011, he worked as an Assistant Lecturer faculty member at AASTMT.

He is currently a Research Assistant at UoM working under the supervision of Prof. Alfa towards the Ph.D. degree in the area of radio resource management over high altitude platforms. He is also currently collaborating in a research project in the domain of using optimization techniques for design purposes in wireless sensor networks.

His research interests mainly include using optimization techniques for radio resource allocation, admission control, multicasting, video traffic scheduling, routing, MAC design, power control and interference management, over different wireless systems like heterogeneous cellular networks, cognitive radio, high altitude platforms, satellites, and wireless sensor networks.

Attahiru S. Alfa is a Professor of telecommunication systems at the Department of Electrical and Computer Engineering, University of Manitoba, and also a SARChI Chair Professor at the Department of Electrical, Electronic and Computer Engineering, University of Pretoria. His research covers, but is not limited to, the following areas: performance analysis and resource allocation in telecommunication systems, modeling of communication networks, queueing theory, optimization, analysis of cognitive radio networks, modeling and analysis of wireless sensor networks, developing efficient decoding algorithms for LDPC codes, channel modeling, traffic estimation for the Internet, and cross layer analysis. He also works in the application of queueing theory to other areas such as transportation systems, manufacturing systems and healthcare systems. He was NSERC Chair for tele-traffic from 2004 to 2012. He has carried out applied research for Nortel Networks, Bell-Northern Research, TRLabs (now TRTech), Bell Canada, Winnipeg Regional Health Authority, Motor-coach Industries, and several other industries. He teaches courses on telecommunication networks, queueing theory, optimization, and network theory. He has authored a book, *Queueing Theory for Telecommunications: Discrete Time Modelling of a Single Node System* (Springer 2010). He is currently working on another book, *Applied Discrete Time Queueing Theory* (Springer 2015).