Relay Selection and Resource Allocation for Multi-User Cooperative OFDMA Networks

Md Shamsul Alam, *Student Member, IEEE,* Jon W. Mark, *Life Fellow, IEEE,*
and Xuemin (Sherman) Shen, *Fellow, IEEE*

***Abstract*—The resource allocation problem is investigated for relay-based multi-user cooperative Orthogonal Frequency Division Multiple Access (OFDMA) uplink system, considering heterogeneous services. A quality of service (QoS) aware optimal relay selection, power allocation and subcarrier assignment scheme under a total power constraint is proposed. The relay selection, power allocation and subcarrier assignment problem is formulated as a joint optimization problem with the objective of maximizing the system throughput, which is solved by means of a two level dual decomposition and subgradient method. To further reduce the computational cost, two low-complexity suboptimal schemes are also proposed. The performance of the proposed schemes is demonstrated through computer simulations based on LTE-A network. Numerical results show that the proposed schemes support heterogeneous services while guaranteeing each user's QoS requirements with slight total system throughput degradation.**

***Index Terms*—OFDMA networks, cooperative relaying, relay selection, resource allocation, joint optimization, QoS, LTE-A.**

I. INTRODUCTION

WITH the rapid development in broadband wireless access technology and explosive growth in demand for new wireless cellular services, it is expected that the next generation cellular network will support a wide variety of communication services with diverse QoS requirements. According to the performance and technical requirements for the 4G networks defined by the International Telecommunication Union (ITU), future International Mobile Telecommunications (IMT)-Advanced mobile system will support very high peak data rates for mobile users, up to 1 Gb/s in static and pedestrian environments, and up to 100 Mb/s in high-speed mobile environment [2]. In order to meet this increasing demand, high-spectral-efficiency schemes are required in conjunction with aggressive resource reuse strategies to ensure prudent use of the scarce radio resources. OFDMA is accepted as the most appropriate air interface for the 4G networks due to it's inherent ability to combat frequency-selective fading and higher spectral efficiency. In OFDMA, multi-user diversity (MUD) can be achieved by allowing subcarriers to be shared among multiple users. Additionally, different number of subcarriers can be allocated to users depending on their QoS requirements. OFDMA is popularly used in 4G wireless systems of broadband communications such as Third Generation Partnership Project (3GPP) Long Term Evolution (LTE) [3], LTE-Advanced [4], Worldwide Interoperability for Microwave Access (WiMAX) [5], and so on.

One of the main challenges facing the 4G networking community is the provision of high throughput for mobiles at the cell edge. Users at the cell edge often suffer from bad channel conditions. Moreover, in an urban environment, shadowing by various obstacles can degrade the signal quality significantly. Cooperative relaying is a very promising solution to tackle this problem as it provides throughput gains as well as coverage extension [6]. Combining OFDMA and cooperative relaying assures high throughput requirements, particularly for users at the cell edge. Additionally, relaying is considered as a cost effective throughput enhancement in both IEEE 802.16j and LTE-A standards. However, to fully exploit the benefits of relaying in 4G networks, efficient relay selection and resource allocation are crucial in multi-user and multi-relay environment.

Resource allocation in OFDMA-based cellular networks without relay has been studied [7]–[9]. Choosing the best relay and allocating the resources in an OFDMA relay network with single user and multiple relays are straightforward and have been well investigated [10], [11]. In the presence of multiple users and multiple relays, relay selection and resource allocation are complicated due to the interactions among the users. An isolated relay assignment and power allocation scheme for cooperative networks considering homogeneous traffic is proposed in [12]. A heuristic algorithm is presented to find a near optimal relay assignment and power allocation where each user is supported by a single relay. However, this scheme can not achieve the optimal solution because of the isolated design approach and the relay selection criterion only based on the maximum allocated power.

There have been numerous research works considering the downlink OFDMA systems [13]–[15]. However, resource allocation schemes designed for the downlink may not be applicable for the uplink due to the distributive nature of power constraints [16]. A multi-user joint distributed resource allocation scheme for uplink cooperative OFDMA system is proposed in [17]. Using the primal-dual decomposition method, the authors have provided an optimal solution and

Manuscript received May 8, 2012; revised November 22, 2012; accepted January 31, 2013. The associate editor coordinating the review of this paper and approving it for publication was Y. Chen.

The authors are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada, (e-mail: {ms3alam, jwmark, sshen}@uwaterloo.ca).

This work has been supported by the Natural Science and Engineering Research Council (NSERC) of Canada under Grant No. RGPIN7779. Part of the paper has been presented in the International Conference on Communications (ICC), 2012, Ottawa, Canada, June 2012 [1].

Digital Object Identifier 10.1109/TWC.2013.032113.120652

distributed implementation. However, relay selection is performed for each user and all subcarriers allocated to that user use the same relay, and each user's QoS requirement is not considered in the joint design. A most recent work [18] proposed a low complexity suboptimal algorithm for subcarrier assignment, power allocation and partner selection for amplify-forward cooperative multicarrier systems. However, homogeneous users with same service and demand are considered. Resource allocation supporting each user's QoS requirements has been considered in several works [16], [19], [20]. In [20], a rate adaptive joint subcarrier and power allocation algorithm under interference and QoS constraints is proposed for cooperative OFDMA based broadband wireless access networks. However, the problem is solved heuristically. A cross layer approach for uplink OFDMA based cellular networks supporting heterogeneous services is introduced in [16]. The authors formulated two different optimization problems to support two types of uplink flows and determined cross-layer trade-off between uplink service rate and power consumption of users. Finally, they solved the problem using dual decomposition.

In this paper, we investigate the joint relay selection and resource allocation problem for the uplink OFDMA-based system. We develop both optimal and suboptimal schemes for relay selection, subcarrier assignment and power allocation with fixed relays, considering service differentiation. The resource allocation problem is formulated as a maximization of the total system throughput by satisfying the individual users' QoS requirements subject to a total power constraint. We consider two types of users, Guaranteed Bit Rate (GBR) users and Aggregate Maximum Bit Rate (AMBR) users. The users are differentiated on the basis of minimum required data rate. GBR users have a specific rate requirement (e.g., real-time gaming) and AMBR users have a flexible service rate (e.g., best-effort and non-real-time service). By relaxing the integer constraints, we derive an optimal solution for this relaxed problem via a two level dual decomposition with reduced computational complexity. We also present two suboptimal schemes based on equal power allocation, with and without power refinement to reduce computational complexity. Numerical results reveal that our proposed schemes significantly outperforms the traditional unconstrained scheme [21] in terms of both services support and QoS satisfaction.

The remainder of this paper is organized as follows. Section II introduces the system model. The problem formulation and analytical framework for the optimal solution are presented in Section III. Suboptimal schemes based on equal power allocation with power refinement are described in Section IV. The computational complexity is discussed in Section V. Numerical results are shown in Section VI, followed by concluding remarks in Section VII.

II. SYSTEM MODEL

Consider a single cell relay enhanced OFDMA-based uplink system with K users (UE) $(1 \leq k \leq K)$ and N fixed relays $(1 \leq n \leq N)$, where relays are shared by all users. The cell is divided into two ring shaped boundary regions and users are distributed between inner and outer boundaries. The reason is that the users located between inner boundary and outer boundary may require relays in most cases due to heavy blockage and long distance transmission [22]. Users located inside the inner boundaries are not considered because they do not require relays in most cases due to good channel condition since they are closer to the eNodeB. Resource allocation for these users may be done separately with simple algorithm. The distance of the relays from the base station (eNodeB) is δR and the relay's angle relative to the base station is uniformly distributed in $[0, 2\pi]$, where R is the radius of the cell and δ is the distance factor. The cell spectrum is divided into subbands, each supported by a subcarrier. The subcarriers are grouped into resource blocks (RBs). The total number of subcarriers used in the system is M $(1 \leq m \leq M)$. The transmit power of the kth user in the mth subcarrier is $P_{s,k}^{m}$, and the transmit power of the nth relay in the mth subcarrier is $P_{r,n}^{m}$. Assume that each node is equipped with a single antenna and the relays operate in a half duplex mode. The broadband channel is assumed to be frequency-selective Rayleigh fading and the destination node (eNodeB) has perfect channel state information (CSI) of all links. The noise variances of the source-to-relay (SR) links, relay-to-destination (RD) links and source-to-destination (SD) links per subcarrier are denoted by $\sigma_{k,n}^2$, $\sigma_{n,D}^2$, and $\sigma_{k,D}^2$, respectively. The system model is shown in Fig. 1. The network model described here can be used to model the uplink system of relay-based LTE-A[1] and IEEE 802.16j networks.

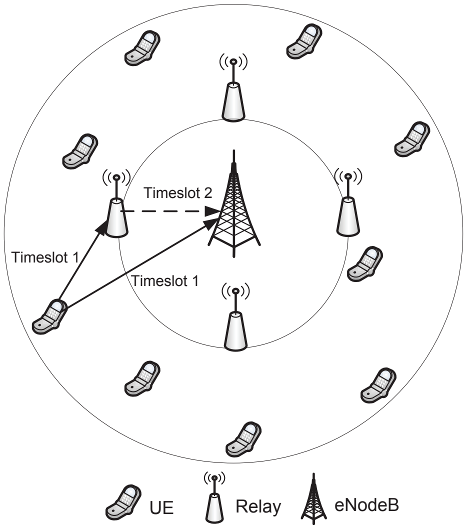

Fig. 1. System model.

There are two types of users: user class κ_1, the GBR users, which have specific rate requirements (called rate constrained (RC) user) and the AMBR users under the user class κ_2, which have a flexible service rate requirements. The traffic class of a user is determined based on the applications. The

---

[1] LTE-A networks adopt Single Carrier Frequency Division Multiple Access (SC-FDMA) in it's uplink, considering the power consumption issue of mobile handsets. Similar to OFDMA downlink, the uplink supports multiple users simultaneously. One prominent advantage of SC-FDMA over OFDMA is that SC-FDMA has significantly lower peak-to-average power ratio (PAPR) [3]. Our network model can be used in LTE-A uplink system under the assumption that subcarriers are not constrained to be consecutive or equidistantly distributed for a higher degree of freedom as considered in [23].

minimum QoS requirement of the kth user is denoted by Q_k. Based on QoS requirement, a user can transmit directly to the destination or transmit using cooperative communication. In cooperative scenario, the communication between the user and the eNodeB is carried out in two phases. In the first phase, the user transmits to the eNodeB which is overheard by the selected relay as well. In the second phase, the selected relay forwards to the eNodeB using the regenerate-and-forward cooperative protocol. The data received in both time slots are combined together by the eNodeB using maximal ratio combining (MRC). The achievable rate in bits/sec/Hz for the regenerate-and-forward scheme for the kth user in the mth subcarrier when the nth relay is selected is given by

$$R_{k,n}^{m} = \begin{cases} \frac{1}{2}\min\left[\log_2(1+P_{s,k}^{m}\alpha_{k,n}^{m})\right., & \\ \left.\log_2(1+P_{s,k}^{m}\alpha_{k,D}^{m}+P_{r,n}^{m}\alpha_{n,D}^{m})\right], & \text{cooperative mode} \\ \log_2(1+P_{s,k}^{m}\alpha_{k,D}^{m}), & \text{non-cooperative mode} \end{cases} \tag{1}$$

where $\alpha_{k,D}^{m} = \frac{\left|h_{k,D}^{m}\right|^2}{\sigma_{k,D}^2}$, $\alpha_{k,n}^{m} = \frac{\left|h_{k,n}^{m}\right|^2}{\sigma_{k,n}^2}$ and $\alpha_{n,D}^{m} = \frac{\left|h_{n,D}^{m}\right|^2}{\sigma_{n,D}^2}$ and $\left|h_{k,D}^{m}\right|^2$, $\left|h_{k,n}^{m}\right|^2$ and $\left|h_{n,D}^{m}\right|^2$ are the channel coefficients between the kth user and the destination, the kth user and the nth relay and the nth relay and the destination in the mth subcarrier, respectively.

Consider binary relay selection and subcarrier allocation characterized by the parameter $\rho_{k,n}^{m}$, where $\rho_{k,n}^{m} = 1$ means that relay node n performs as a relay for user k in the mth subcarrier. Otherwise, it is equal to 0. We assume that each user can have only one relay, but each relay can support several users and a subcarrier is allocated to only one source and one relay, so that there is no interference between sources. The same subcarrier will be used by the relay in the second time slot. Even if it is decided that relay will not transmit in the second time slot (i.e. non-cooperative mode), the user is not allowed to use this idle time slot [21].

III. PROBLEM FORMULATION AND SOLUTION APPROACH

Our objective is to maximize the total system throughput subject to a set of constraints. The relay selection and subcarrier assignment constraints are as follows:

$$\sum_{k=1}^{K}\sum_{n=0}^{N}\rho_{k,n}^{m} = 1, \rho_{k,n}^{m} \in \{0,1\}, \forall m \tag{2}$$

where $n = 0$, it means user k utilize subcarrier m in non-cooperative mode. The total power allocated to the mth subcarrier of the kth user in both time slots is $P_{t,k}^{m} = P_{s,k}^{m} + P_{r,n}^{m}$ [10], [24] and the total power constraint can be expressed as

$$\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{n=0}^{N}\rho_{k,n}^{m}P_{t,k}^{m} \le P_T \tag{3}$$

where P_T is the sum of the power available for all users plus relays in the network. Although individual power constraints will lead more accurate power allocation, however, our goal is to maximize the total system throughput subject to a joint total power constraint, considering the simplicity of the problem formulations and lower computational complexity under the sum power constraint. The computational complexity is lower in the studied model since we only need to update one dual variable using subgradient method under the total power constraint compared to updating $K+N$ dual variables simultaneously until all of them are converged when individual power constraints are used. Similar assumptions on the total power constraint are taken in previous studies [10], [24]–[26].

Maximization of the rate in (1) using cooperative communication under total power constraint has advantageous only if $\alpha_{k,n}^{m} > \alpha_{k,D}^{m}$ and $\alpha_{n,D}^{m} > \alpha_{k,D}^{m}$ [25], [27]. First, consider the case when the user to relay channel is weaker (lower Signal-to-Noise Ratio (SNR) due to bad channel condition) than the direct link channel, i.e., $\alpha_{k,n}^{m} < \alpha_{k,D}^{m}$. In such case, from equation (1), any power increment will be more beneficial if allocated to the direct link and the use of relay will not be advantageous. Second, consider the case when the user to relay channel is stronger (higher SNR due to good channel condition) than the direct link channel, .i.e., $\alpha_{k,n}^{m} > \alpha_{k,D}^{m}$. Then two cases may happen: 1) if $\alpha_{k,D}^{m} > \alpha_{n,D}^{m}$, the rate benefit will be greater if the power is allocated to the direct link; and 2) if $\alpha_{k,D}^{m} < \alpha_{n,D}^{m}$, the allocation of power to the relay is better under the constraint of $P_{s,k}^{m}\alpha_{k,D}^{m} + P_{r,n}^{m}\alpha_{n,D}^{m} \le P_{s,k}^{m}\alpha_{k,n}^{m}$. This means that any power increment has to be shared between the user and relay, and the rate will be maximized when the constraint is saturated,.i.e., $P_{s,k}^{m}\alpha_{k,D}^{m} + P_{r,n}^{m}\alpha_{n,D}^{m} = P_{s,k}^{m}\alpha_{k,n}^{m}$. Then the source power allocation is given by

$$P_{s,k}^{m} = \begin{cases} \frac{\alpha_{n,D}^{m}}{\alpha_{k,n}^{m}+\alpha_{n,D}^{m}-\alpha_{k,D}^{m}}P_{t,k}^{m}, & \text{cooperative mode} \\ P_{t,k}^{m}, & \text{non-cooperative mode} \end{cases} \tag{4}$$

and the relay power allocation is given by

$$P_{r,n}^{m} = \begin{cases} \frac{\alpha_{k,n}^{m}-\alpha_{k,D}^{m}}{\alpha_{k,n}^{m}+\alpha_{n,D}^{m}-\alpha_{k,D}^{m}}P_{t,k}^{m}, & \text{cooperative mode} \\ 0, & \text{non-cooperative mode} \end{cases} \tag{5}$$

The computation of the source and relay power can be explained as follows. First, for any subcarrier, when the channel gains are known, the transmission mode can be determined for the given total power on that subcarrier. Second, for the selected transmission mode, the optimal source and relay power are computed. If cooperative mode is selected, $P_{t,k}^{m}$ will be divided in two time slots depending on the channel condition and the optimal source and relay power are given by (4) and (5) [10], [25]. In case of non-cooperative mode, $P_{r,n}^{m} = 0$, and $P_{s,k}^{m} = P_{t,k}^{m}$ from (4). For the non-cooperative transmission mode, there may be two scenarios: transmission can be held in two time slots by dividing the total power, $P_{t,k}^{m}$ in two time slots, or use only one time slot with power $P_{t,k}^{m}$. In this work, we assume that the user only transmits in the first time slot using the total power, $P_{t,k}^{m}$ when non-cooperative transmission mode is selected [21]. Thus, substituting (4) and (5) into (1), the rate expression can be unified as

$$R_{k,n}^{m} = \frac{1}{2}[\log_2(1+P_{t,k}^{m}\alpha_{k,eq}^{m})] \tag{6}$$

where $\alpha_{k,eq}^{m}$ is the equivalent channel gain given by

$$\alpha_{k,eq}^{m} = \begin{cases} \frac{\alpha_{k,n}^{m}\alpha_{n,D}^{m}}{\alpha_{k,n}^{m}+\alpha_{n,D}^{m}-\alpha_{k,D}^{m}}, & \text{cooperative mode} \\ \alpha_{k,D}^{m}, & \text{non-cooperative mode.} \end{cases} \tag{7}$$

The total achievable rate of the kth user for all subcarriers allocated to the kth user is given by

$$R_k = \sum_{m=1}^{M} \sum_{n=0}^{N} \rho_{k,n}^{m} R_{k,n}^{m}. \tag{8}$$

We formulate the joint resource allocation and relay selection problem subject to a minimum data rate constraint for each GBR user. The optimization problem can be formulated as

$$\begin{aligned} (P1) \quad & \underset{\rho, P_t}{\text{maximize}} \quad \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{n=0}^{N} \rho_{k,n}^{m} R_{k,n}^{m} \\ & \text{subject to} \quad c1: \rho_{k,n}^{m} \in \{0,1\}, \forall k, m, n \\ & \qquad c2: \sum_{k=1}^{K} \sum_{n=0}^{N} \rho_{k,n}^{m} = 1, \forall m \\ & \qquad c3: R_k \geq Q_k, \forall k \in \kappa_1 \\ & \qquad c4: \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{n=0}^{N} \rho_{k,n}^{m} P_{t,k}^{m} \leq P_T \\ & \qquad c5: P_{t,k}^{m} \geq 0, \forall k, m, n \end{aligned} \tag{9}$$

where constraints $c1$ and $c2$ represent the relay selection and subcarrier allocation and indicate that each user can have one relay to cooperate and can utilize multiple subcarriers to transmit; however, a subcarrier can not be shared by different users. Constraint $c3$ applies minimum QoS requirements for the GBR users in terms of data rate requirement. Finally, the source and the relay power allocation are constrained by $c4$ and $c5$.

The optimization problem in (9) is a mixed integer nonlinear programming (MINP) problem. One challenging aspect of this problem in the context of OFDMA uplink is the discrete nature of subcarrier assignment, which, when coupled with QoS constraint, makes the problem even harder to solve. Therefore, finding the optimal solution for this non-convex problem requires searching through all the possible user, relay and subcarrier allocations, which is prohibitively complex to employ in large system. However, to make the problem tractable, we relax the integer constraints, $\rho_{k,n}^{m}$ to take any real value between 0 and 1 via time-sharing condition which allows time sharing of each subcarrier. The duality gap of any optimization problem satisfying the time sharing condition is negligible as the number of subcarriers becomes sufficiently large [28]. Since our optimization problem obviously satisfies the time-sharing condition, it can be solved by using the dual method and the solution is optimal [16], [28].

A. Dual Problem

The Lagrangian function of problem in (9) can be written as

$$\begin{aligned} & L(\rho, P_t, \lambda, \mu) \\ &= \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{n=0}^{N} \rho_{k,n}^{m} R_{k,n}^{m} + \sum_{k \in \kappa 1} \lambda_k \Big(\sum_{m=1}^{M} \sum_{n=0}^{N} \rho_{k,n}^{m} R_{k,n}^{m} \\ & \quad - Q_k \Big) + \mu \Big(P_T - \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{n=0}^{N} \rho_{k,n}^{m} P_{t,k}^{m} \Big) \\ &= \sum_{m=1}^{M} \Bigg[\sum_{k=1}^{K} \sum_{n=0}^{N} \rho_{k,n}^{m} R_{k,n}^{m} + \sum_{k \in \kappa 1} \lambda_k \sum_{n=0}^{N} \rho_{k,n}^{m} R_{k,n}^{m} \\ & \quad - \mu \sum_{k=1}^{K} \sum_{n=0}^{N} \rho_{k,n}^{m} P_{t,k}^{m} \Bigg] - \sum_{k \in \kappa 1} \lambda_k Q_k + \mu P_T \end{aligned} \tag{10}$$

where $\lambda = [\lambda_1, \lambda_2, ...\lambda_{\kappa 1}]^T$ is the vector of the dual variables associated with the individual QoS constraints and μ is the dual variable for the power constraint. The Lagrangian dual function can therefore be written as

$$g(\lambda, \mu) = \begin{cases} \underset{\rho, P_t}{\max} & L(\rho, P_t, \lambda, \mu) \\ \text{s.t.} & \sum_{k=1}^{K} \sum_{n=0}^{N} \rho_{k,n}^{m} = 1, \forall m \\ & 0 \geq \rho_{k,n}^{m} \leq 1, P_{t,k}^{m} \geq 0. \end{cases} \tag{11}$$

Then the dual optimization problem is given by

$$\min_{\lambda, \mu \succeq 0} g(\lambda, \mu). \tag{12}$$

The coupling between subcarriers via Lagrangian relaxation can be removed and (11) can be decomposed into M subproblems at each subcarrier, which can be solved independently given λ, μ with low complexity. The subproblem at each subcarrier is given by

$$\begin{aligned} & \max_{\rho, P_t} \; L_m(\rho^m, P_t^m) \\ &= \max_{\rho, P_t} \sum_{k=1}^{K} \sum_{n=0}^{N} \rho_{k,n}^{m} R_{k,n}^{m} + \sum_{k \in \kappa 1} \lambda_k \sum_{n=0}^{N} \rho_{k,n}^{m} R_{k,n}^{m} \\ & \quad - \mu \sum_{k=1}^{K} \sum_{n=0}^{N} \rho_{k,n}^{m} P_{t,k}^{m} \\ & s.t. \quad \sum_{k=1}^{K} \sum_{n=0}^{N} \rho_{k,n}^{m} = 1, 0 \geq \rho_{k,n}^{m} \leq 1, P_{t,k}^{m} \geq 0, \forall k, n \end{aligned} \tag{13}$$

where ρ^m, P_t^m are the vectors of $\rho_{k,n}^{m}, P_{t,k}^{m}$ on the mth subcarrier, respectively. The subproblem can be further decomposed through a second level primal decomposition. The decomposition hierarchy of the dual problem is shown in Fig. 2. Thus, we have two subproblems which will be solved in two phases: optimal power allocation and joint relay selection and subcarrier allocation.

Proposition 1: Considering the convex optimization problem in (12), the subgradients of $g(\lambda, \mu)$ denoted by $\Delta\lambda_k$, and $\Delta\mu$ are given by

$$\Delta\lambda_k = \textstyle\sum_{m=1}^{M} \sum_{n=0}^{N} \rho_{k,n}^{m\star} R_{k,n}^{m\star} - Q_k, \forall k \in \kappa_1$$

$$\Delta\mu = P_T - \textstyle\sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{n=0}^{N} \rho_{k,n}^{m\star} P_{t,k}^{m\star}$$

where $\rho_{k,n}^{m\star}, R_{k,n}^{m\star}$ and $P_{t,k}^{m\star}$ are the optimal solution of the dual objective function in (12).

The proof of proposition 1 is given in the Appendix.

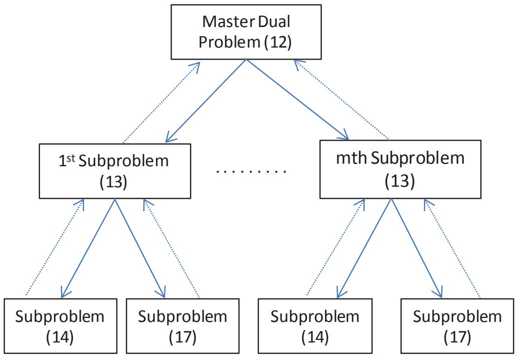

Fig. 2. Hierarchy of the decomposed dual problem.

B. Optimal Power Allocation for a Given Relay Assignment and Subcarrier Allocation

Let subcarrier m be allocated to user k and relay n in a frame of transmission time and $\rho^m_{k,n} = 1$. Then optimal power allocation over this subcarrier and relay assignment can be determined by solving the following problem

$$\begin{aligned} \max_{P^m_{t,k}} \quad & L_m, \forall k, n \\ \text{s.t.} \quad & P^m_{t,k} \geq 0. \end{aligned} \tag{14}$$

Substituting (6) into (14) and differentiating L with respect to $P^m_{t,k}$ we have

$$\frac{\partial L}{\partial P^m_{t,k}} = \frac{(1+\bar{\lambda}_k)\alpha^m_{k,eq}}{2ln(2)(1+P^m_{t,k}\alpha^m_{k,eq})} - \mu \tag{15}$$

where

$$\bar{\lambda}_k = \begin{cases} \lambda_k, \forall k \in \kappa_1 \\ 0, \text{Otherwise.} \end{cases}$$

Applying the Karush-Kuhn-Tucker (KKT) [29] condition, we can deduce the optimal power allocation as follows:

$$P^{m\,*}_{t,k} = \left[\frac{1+\bar{\lambda}_k}{2\mu\ln(2)} - \frac{1}{\alpha^m_{k,eq}}\right]^+ \tag{16}$$

where $[x]^+ = max\,[x, 0]$.

C. Joint Optimal Relay Selection and Subcarrier Allocation

By eliminating the power variables in (14) and then substituting into (10), we have an alternative expression of the dual function as

$$\begin{aligned} g(\lambda,\mu) \quad = \quad & \max_{\rho} \sum_{m=1}^{M}\sum_{k=1}^{K}\sum_{n=0}^{N} \rho^m_{k,n} H^m_{k,n}(\lambda,\mu) \qquad (17) \\ & - \sum_{k\in\ \kappa 1} \lambda_k Q_k + \mu P_T \\ s.t. \quad & \sum_{k=1}^{K}\sum_{n=0}^{N} \rho^m_{k,n} = 1, \forall m, 0 \geq \rho^m_{k,n} \leq 1 \end{aligned}$$

where the function $H^m_{k,n}(\lambda,\mu)$ is defined as follows

$$H^m_{k,n} = \frac{1}{2}(1+\bar{\lambda}_k)[\log_2(1+P^{m\,*}_{t,k}\alpha^m_{k,eq})] - \mu P^{m\,*}_{t,k}. \tag{18}$$

An intuitive explanation for each term in (18) is as follows. The first term can be viewed as the rate obtained by selecting subcarrier m by user k and relay n and the second term is the price for the power consumption. Therefore, $H^m_{k,n}$ can be interpreted as the gain of transmitting over subcarrier m by user k and relay n and $\mathbf{H} = [H^m_{k,n}]$ can be represented as a $K \times N$ profit matrix at each subcarrier m. In other words, the profit matrix $\mathbf{H}$ is different for different value of m. The objective function in (17) can be maximized by picking exactly one element of matrix $\mathbf{H}$ for each subcarrier so that the sum of profit is as large as possible. Finally, optimal relay selection and subcarrier allocation should be the one having the maximum value of $H^m_{k,n}(\lambda,\mu)$ in (18) and is given by

$$\rho^m_{k,n} = \begin{cases} 1, (n^\star, k^\star) = arg \ \max\limits_{n,k} \ H^m_{k,n} \\ 0, \text{otherwise.} \end{cases} \tag{19}$$

In the operation, first, the power allocation for each subcarrier using both transmission modes is computed using (16). Then, these power allocation values are used in (18) to compute $H^m_{k,n}$. After that, for each subcarrier, the user and relay pair is determined using (19) that gives the largest $H^m_{k,n}$. Non-cooperative mode is the case that no relay is selected, i.e., $n = 0$.

D. Variable Update

Since a dual function is always convex by definition, subgradient method can be used to minimize $g(\lambda,\mu)$. Dual variables λ and μ are updated in parallel as follows

$$\begin{aligned} \lambda_k(t+1) &= \left[\lambda_k(t) + \eta(t)\left(Q_k - \sum_{n=0}^{N}\sum_{m=1}^{M} \rho^m_{k,n}(t) R^m_{k,n}(t)\right)\right]^+ \\ \mu(t+1) &= \left[\mu(t) + \theta(t)\left(\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{n=0}^{N} \rho^m_{k,n}(t) P^m_{t,k}(t) - P_T\right)\right]^+ \end{aligned} \tag{20}$$

where $\eta(t)$ and $\theta(t)$ are diminishing stepsizes and t is the iteration index. The subgradient method above is guaranteed to converge to the optimal dual variables if the stepsizes are chosen following the diminishing stepsize policy [29]. Based on the mathematical formulations and derivations, the optimal relay selection, subcarrier assignment and power allocation can be computed algorithmically. The pseudocode of the proposed optimal scheme is outlined in Algorithm 1.

IV. SUBOPTIMAL SCHEMES

The computational complexity of the proposed optimal scheme may still be too high for practical implementation. In this section, we present two suboptimal schemes which have lower computational cost compared to the optimal one.

A. Equal Power Allocation (EPA) Scheme

In this scheme, we determine relay selection and subcarrier allocation assuming that the power is equally distributed over all subcarriers. First, relay selection and subcarrier allocation are performed for the GBR users in two steps considering that AMBR users are absent. In step 1, to ensure fairness among the users, we select the user whose current achievable rate is

Algorithm 1 Pseudocode for the proposed optimal scheme

```
Initialize $\lambda_k$ and $\mu$.
for $m = 1 \rightarrow M$ do
    for $k = 1 \rightarrow K$ do
        for $n = 1 \rightarrow N$ do
            Calculate optimal power using (16).
            Calculate $H_{k,n}^{m}(\lambda, \mu)$ using (18).
        end for
    end for
    Find optimal $(n^\star, k^\star)$ according to (19).
    Allocate subcarrier $m$ to $(n^\star, k^\star)$.
end for
for $k = 1 \rightarrow K$ do
    update $\lambda_k$ using (20).
end for
Update $\mu$ using (20).
Repeat above steps until convergence.
```

the farthest away from it's minimum rate requirement. In step 2, for the selected user, we choose the subcarrier and relay that maximize the transmission rates $R_{k,n}^{m}$, rather than the metrics $H_{k,n}^{m}$ defined in (18). Steps 1 and 2 are repeated until all users are satisfied or the number of unassigned subacarriers are zero. Then the remaining subcarriers and power are distributed among the AMBR users to maximize the sum rate. In this case, to exploit multi-user diversity, subcarriers are allocated to the user and relay pair who can utilize the channel the best. Let S_k be the set of subcarriers assigned to user k and A be the set of unassigned subcarriers. The pseudocode of the EPA scheme is presented in Algorithm 2.

Algorithm 2 Pseudocode for EPA scheme

```
Initialization: set $R_k = 0$, $S_k = \phi$, $\forall k$ and $A = \{1, 2, ....M\}$.
STEP 1: GBR users
while $A \neq \phi$ and $R_k < R_Q$ for all $k \in \kappa_1$ do
    Select user $k^\star = arg \quad \max_{k \in \kappa_1}(R_Q - R_k)$.
    For the found $k^\star$, find $(n^\star, m^\star) = arg \quad \max_{n,m} \quad R_{k,n}^{m}$.
    Assign the subcarrier $m^\star$ to $(n^\star, k^\star)$ .
    Update $S_{k^\star} = S_{k^\star} \cup m^\star$, $A = A - m^\star$ and $R_{k^\star} = R_{k^\star} + R_{k^\star,n^\star}^{m^\star}$.
end while
STEP 2: AMBR users
while $A \neq \phi$ do
    Find $(n^\star, k^\star, m^\star) = arg \quad \max_{n,k \in \kappa 2,m} \quad R_{k,n}^{m}$.
    Update $S_{k^\star} = S_{k^\star} \cup m^\star$, $A = A - m^\star$ and $R_{k^\star} = R_{k^\star} + R_{k^\star,n^\star}^{m^\star}$.
end while
```

B. Equal Power Allocation with Refinement (EPAR) Scheme

In this suboptimal scheme, we use power refinement after relay selection and subcarrier assignment with equal power distribution. First, relay selection and subcarrier assignment are done for the GBR users considering equal power distribution. Then the power distribution for each GBR user is adjusted individually using analytical solution described in Section IV-C. The objective of the power refinement is to optimize the power while maintaining the basic transmission rate. Subcarrier adjustment is performed after the power refinement. In the subcarrier adjustment substep, some GBR users release additional subcarriers which are over allocated by the EPA scheme and can be used for AMBR users. At the end, the remaining subcarriers and power are allocated to the AMBR users and power refinement is also done by using the analytical solution presented in Section IV-C. Since the subcarrier assignment for the GBR users in the first step is obtained by considering equal power distribution, this scheme is suboptimal. The computational complexity of the power refinement process is much smaller than that of the dual problem because we already have relay selection and subcarrier assignment. The pseudocode of the EPAR scheme is presented in Algorithm 3.

Algorithm 3 Pseudocode for EPAR scheme

```
Initialization: set $R_k = 0$, $S_k = \phi$, $\forall k$ and $A = \{1, 2, ....M\}$.
STEP 1: GBR users
while $A \neq \phi$ and $R_k < R_Q$ for all $k \in \kappa_1$ do
    Select user $k^\star = arg \quad \max_{k \in \kappa_1}(R_Q - R_k)$.
    For the found $k^\star$, find $(n^\star, m^\star) = arg \quad \max_{n,m} \quad R_{k,n}^{m}$.
    Assign the subcarrier $m^\star$ to $(n^\star, k^\star)$ .
    Update $S_{k^\star} = S_{k^\star} \cup m^\star$, $A = A - m^\star$ and $R_{k^\star} = R_{k^\star} + R_{k^\star,n^\star}^{m^\star}$.
end while
STEP 2: Power Refinement for GBR users
Refine power using power refinement method stated in Section IV-C.
STEP 3: Subcarrier Adjustment
for each user $k \in \kappa_1$ do
    while $R_k > R_Q$ do
        Find $m' = arg \quad \min_{m \in S_k}(R_{k,n}^{m})$
        if $(R_k - R_{k,n}^{m'}) \geq R_Q$ then
            Update $S_k = S_k - m'$, $A = A \cup m'$, $R_k = R_k - R_{k,n}^{m'}$.
        end if
    end while
end for
STEP 4: AMBR users
while $A \neq \phi$ do
    Find $(n^\star, k^\star, m^\star) = arg \quad \max_{n,k \in \kappa 2,m} \quad R_{k,n}^{m}$.
    Update $S_{k^\star} = S_{k^\star} \cup m^\star$, $A = A - m^\star$ and $R_{k^\star} = R_{k^\star} + R_{k^\star,n^\star}^{m^\star}$.
end while
STEP 5: Power Refinement for AMBR users
Refine power using power refinement method stated in Section IV-C.
```

C. Power Refinement: Method 1

In power refinement method 1, we optimize the power while maximizing the throughput for a given subcarrier and relay assignment, and guaranteeing the minimum rate requirements for each GBR user. In other words, it determines optimal power for a given subcarrier allocation and relay assignment while maximizing the total system throughput and maintaining the basic transmission rates. First, the relay selection and subcarrier allocation are obtained using equal power distribution similar to the EPA scheme. Then, the optimal power allocation is performed on this subcarrier-relay assignment. The optimization problem can be expressed as follows:

$$(P2) \quad \max_{P_t} \quad \sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{n=0}^{N}\rho_{k,n}^{m}R_{k,n}^{m} \tag{21}$$

subject to constraints $c3$, $c4$ and $c5$ in (9).

Optimization problem $P2$ is a convex optimization problem which can be solved similarly as the optimal power allocation approach described in Section III-B. The key difference with the optimal scheme is that there is no need to optimize the subcarrier assignment and relay selection and the computational complexity of the power refinement process is far smaller than that of solving the dual problem in (10). The pseudocode of the power refinement method 1 is outlined in Algorithm 4.

Algorithm 4 Pseudocode of the proposed power refinement method 1

```
Initialize $\lambda_k$ and $\mu$.
STEP 1: Power Refinement for GBR users
for each user $k \in \kappa_1$ do
    for each subcarrier $m \in S_k$ do
        Calculate optimal power using (16).
    end for
    Update $\lambda_k$ using (20).
end for
Update $\mu$ using (20).
Repeat above steps until convergence.
STEP 2: Power Refinement for AMBR users
for each user $k \in \kappa_2$ do
    for each subcarrier $m \in S_k$ do
        Calculate optimal $P_{t,k}^{m\,*} = \left[\frac{1}{2\mu\ln(2)} - \frac{1}{\alpha_{k,eq}^{m}}\right]^{+}$.
    end for
end for
Update $\mu$ using (20).
Repeat above steps until convergence.
```

D. Power Refinement: Method 2

The previous method still needs to update the Lagrangian multipliers λ and μ to meet the rate and power constraints, respectively. However, to further reduce the computational complexity, we assume that $x_k = \frac{1+\bar{\lambda}_k}{2\mu\ln 2}, \forall k \in \kappa_1$; then from (16) we have

$$x_k = P_{t,k}^{m\,*} + \frac{1}{\alpha_{k,eq}^{m}} \tag{22}$$

where $P_{t,k}^{m\,*}$ can be represented as $P_{t,k}^{m} = \frac{P_T}{M}$ under equal power distribution. Rewriting (22) as

$$|S_k|x_k = \sum_{m\in S_k}\left[P_{t,k}^{m} + \frac{1}{\alpha_{k,eq}^{m}}\right], \forall k \in \kappa_1 \tag{23}$$

where $|S_k|$ is the cardinality of S_k. Finally, x_k can be deduced as

$$x_k = \frac{1}{|S_k|}\left[P_{t,k} + \sum_{m\in S_k}\frac{1}{\alpha_{k,eq}^{m}}\right], \forall k \in \kappa_1 \tag{24}$$

where $P_{t,k} = |S_k|P_{t,k}^{m}$. Substituting x_k into (16), the power allocation refinement can be obtained for GBR users.

For AMBR users, setting $\bar{\lambda}_k = 0$ in (16), we get the optimal power allocation for AMBR users as

$$P_{t,k}^{m\,*} = \left[\frac{1}{2\mu\ln 2} - \frac{1}{\alpha_{k,eq}^{m}}\right]^{+}, \forall k \in \kappa_2. \tag{25}$$

Letting $y_k = \frac{1}{2\mu\ln 2}, \forall k \in \kappa_2$, y_k can be deduced as

$$y_k = \frac{1}{|S_k|}\left[P_{t,k} + \sum_{m\in S_k}\frac{1}{\alpha_{k,eq}^{m}}\right], \forall k \in \kappa_2 \tag{26}$$

where $P_{t,k}$ is given by

$$P_{t,k} = \frac{|S_k|(P_T - \sum_{k\in\kappa_1}\sum_{m=1}^{M}P_{t,k}^{m})}{M - \sum_{k\in\kappa_1}|S_k|}. \tag{27}$$

Finally, the power refinement for the AMBR users is obtained by substituting (26) into (25). The pseudocode of the power refinement method 2 is outline in Algorithm 5.

Algorithm 5 Pseudocode of the proposed power refinement method 2

```
STEP 1: Power Refinement for GBR users
for each user $k \in \kappa_1$ do
    for each subcarrier $m \in S_k$ do
        Calculate $P_{t,k}^{m} = \frac{P_T}{M}$.
        Calculate $x_k$ using (24).
        Calculate optimal power using (16).
    end for
end for
STEP 2: Power Refinement for AMBR users
for each user $k \in \kappa_2$ do
    for each subcarrier $m \in S_k$ do
        Calculate $P_{t,k}$ using (27).
        Calculate $y_k$ using (26).
        Calculate optimal power using (25).
    end for
end for
```

V. COMPLEXITY ANALYSIS

The computational complexity of the proposed optimal scheme is mainly determined by the complexity of solving the dual problem. The total number of computations needed to perform relay selection is $K(N+1)$ and M allocations are required for all subcarriers. Therefore, the complexity at each iteration is $O(MKN)$. The complexity of the subgradient method is polynomial in the number of dual variables. With the total power constraint, there are $\kappa_1 + 1$ dual variables and the overall complexity is $O(|\kappa_1|^2 MKN)$. The complexity of the whole scheme is polynomial, which is significantly lower than employing the exhaustive search solution to the master primal problem because the number of subcarrier assignment policies increases exponentially with M.

The complexity of the EPA scheme can be analyzed as follows. The complexity of allocating subcarriers to the GBR users is $O(|\kappa_1|MN)$ and the complexity of allocating the remaining subcarriers to the AMBR users is $O(|\kappa_2|(M - |\kappa_1|)N)$. So, the overall complexity of the EPA scheme is $O(|\kappa_1|MN)$ + $O(|\kappa_2|(M - |\kappa_1|)N)$.

The complexity of the EPAR scheme with power refinement method 1 depends on the convergence of the subgradient method. The complexity of allocating subcarriers to both the GBR and AMBR users using the EPAR scheme is same as the EPA scheme. The complexity of the power refinement using method 1 depends on the number of iterations required for the subgradient methods to converge. There are $\kappa_1 + 1$ dual variables and the overall complexity for GBR users is $O(|\kappa_1|^2 M)$. Since there is only one dual variables, the complexity of power refinement method 1 for AMBR users is $O(M - |\kappa_1|)$. Thus, the overall complexity of the EPAR scheme using power refinement method 1 is $O(|\kappa_1|MN)$ + $O(|\kappa_1|^2 M)$+ $O(|\kappa_2|(M - |\kappa_1|)N)$ + $O(M - |\kappa_1|)$.

Power refinement method 2 does not require any update of the dual variables. The overall complexity of the EPAR scheme using power refinement method 2 is $O(|\kappa_1|MN)$ + $O(M)$+ $O(|\kappa_2|(M - |\kappa_1|)N)$ + $O(M - |\kappa_1|)$.

VI. NUMERICAL RESULTS

To evaluate the performance of our schemes, numerical results are generated using a simulation scenario. Consider a single cell LTE-A network with a radius of 1Km, where eNodeB is located at the center of the cell. The fixed relays are located at a radius of δ Km from the eNodeB at equal angular distance where δ varies between 0.2 to 0.8. The relay locations are varied to show the effect of relay locations on the performance. Here, we only consider random variations of the relay distance from the eNodeB as the first step. However, relay placement can be modeled as another optimization problem which is not studied in this paper. The UE locations are randomly generated and uniformly distributed between 0.5 Km to 1 Km from the eNodeB. Half of the users in the system are assumed to be GBR and the other half are AMBR users. The GBR users are selected randomly from the total set of users. These GBR users have different rate requirements based on the applications. We allocate different applications to different users arbitrarily, and they are fixed for the whole simulation. Multipath Rayleigh fading with exponential power delay profile based on ITU pedestrian B model [3] is considered for small scale fading model. The channels for different users in each subcarrier are assumed to be independent. Then the effective channel gain over an RB is deduced from the subcarrier granularity. The 3GPP LTE path loss model with log-normal shadowing of an 8dB standard deviation are assumed. The system parameters are given in Table I. Having the simulation scenarios and all the system parameters, the optimal relay selection, power allocation and subcarrier assignments are evaluated using Algorithm 1. The stepsize for λ and μ is set to 0.01 divided by $\sqrt{IterationNumber}$. Relay selection is performed per RB since RB is the smallest resource unit for the LTE network. The simulation scenario (user locations, selection of the GBR users and the assignment of the applications to the users) is repeated 100 times to get a fair result. The multipath channel components are repeated over 1000 times.

TABLE I
SIMULATION PARAMETERS

Name of the Parameters	Value
Total system bandwidth	5 MHz
Total number of RB	24
Total number of subscribers	288
Number of UEs	24
Number of relays	2, 4, 6, 8
Total power available at UE	23 dBm
Total power available at relay	30 dBm
Noise power spectral density	-174 dBm/Hz
Path loss exponent	3.76

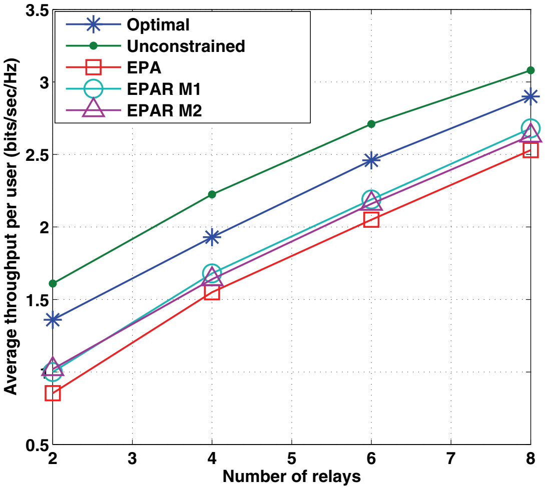

Fig. 3. Average throughput for 24 users with different number of relays, $\delta = 0.5$.

The optimization problem considered in this work may be infeasible due to the rate requirements constraint. This may happen if the channel condition is very bad (low SNR) and/or the available resources are limited to support the minimum rate requirements of the GBR users. In the simulation, we allocate resources as much as possible for the GBR users on those infeasible cases and also consider them when we calculate the average spectral efficiency. Those situations have been further verified and handled by introducing the user satisfaction index (SI). The user satisfaction index (SI) [30] is calculated as $SI = \frac{1}{K}\sum_{k=1}^{K} SI_k$, where $SI_k = min\left(\frac{R_k}{Q_k}, 1\right)$. SI is less than 1 means there are some cases which are infeasible and the minimum rate requirements for some users are not satisfied.

Fig. 3 shows the average throughput per user in bits/sec/Hz for the optimal scheme, suboptimal schemes and traditional unconstrained scheme as a function of the number of relays. The optimal scheme provides slightly lower throughput compared to the unconstrained scheme, because the unconstrained scheme always allocates subcarriers by only considering the channel condition. So, some users have very high rate since most of the subcarriers are allocated to those users, whereas others have very low rate because very few or no subcarriers

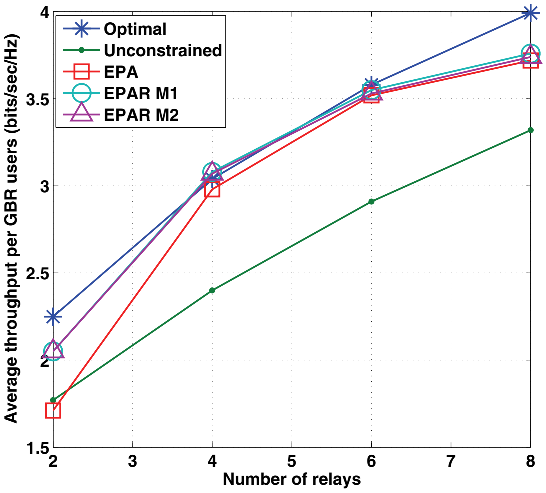

Fig. 4. Average throughput for GBR users as a function of the number of relays, $\delta = 0.5$.

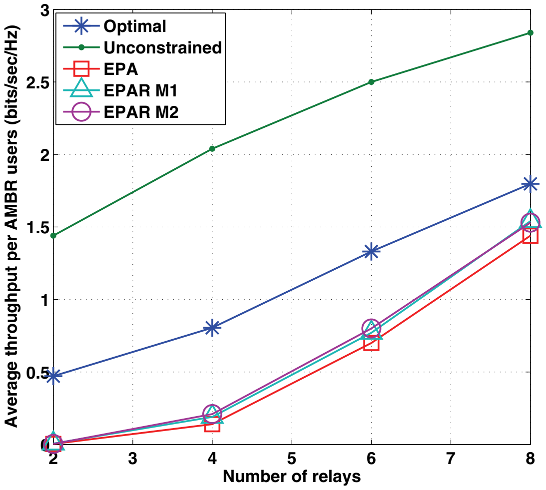

Fig. 5. Average throughput for AMBR users as a function of the number of relays, $\delta = 0.5$.

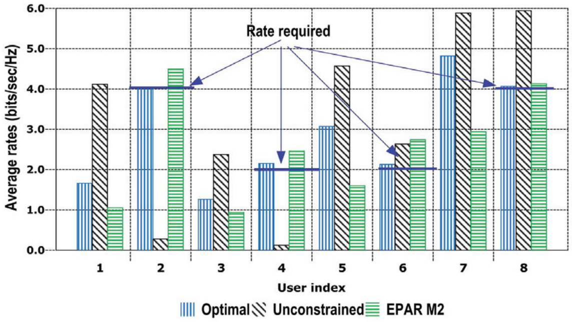

Fig. 6. User achievable rate with 8 relays, rate required for user 2 and user 8 = 4 bits/sec/Hz and rate required for user 4 and user 6 = 2 bits/sec/Hz, $\delta = 0.5$.

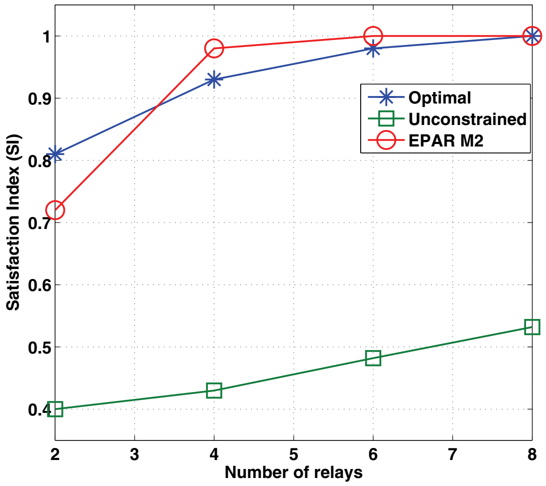

Fig. 7. Average user satisfaction index for GBR users with different number of relays, $\delta = 0.5$.

are allocated to them. On the other hand, the optimal scheme considers both minimum rate requirement as well as channel condition, and distributes the subcarriers to the users based on their minimum rate requirement. So, when we average over the total number of channel realizations, the average throughput is higher for the unconstrained scheme but it violates fairness which is also evident in Fig. 6. However, the performance gap for the optimal scheme reduces with an increase in the number of relays.

It is noted that all suboptimal schemes provide lower throughput compared to the optimal scheme. Because all suboptimal schemes use equal power distribution for relay selection and subcarrier assignment. However, the EPAR scheme with power refinement method 1 (EPAR M1) and the EPAR scheme with power refinement method 2 (EPAR M2) perform well although they have lower computational complexity compared to the optimal one. Both of the power refinement methods have almost similar performance although power refinement method 2 is computationally simpler. The EPAR scheme has higher rates compared to the EPA scheme. This is due to the power refinement and subcarrier adjustment used in the EPAR scheme.

The average throughput for the GBR users for all schemes as a function of the number of relays is shown in Fig. 4. The traditional unconstrained scheme provides the lowest throughput since it does not consider user's minimum rate requirements. The optimal scheme has the highest throughput in all cases. However, all suboptimal schemes exhibit performance close to the optimal scheme as the number of relays increases. The reason is that all suboptimal schemes first allocate subcarriers and power to the GBR users, and when all GBR users are satisfied, the remaining subcarriers and power are then allocated to the AMBR users. So, the reverse characteristic is observed in case of AMBR users for all schemes except the optimal scheme, as shown in Fig. 5. Since the AMBR users have no minimum rate requirements, the unconstrained scheme provides the highest throughput.

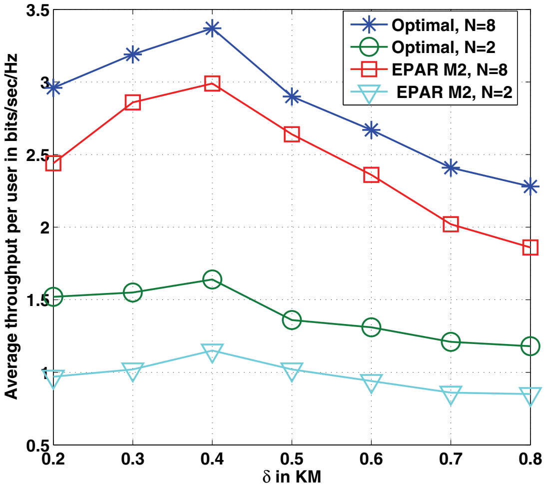

Fig. 8. Average throughput for 24 users with different relay locations.

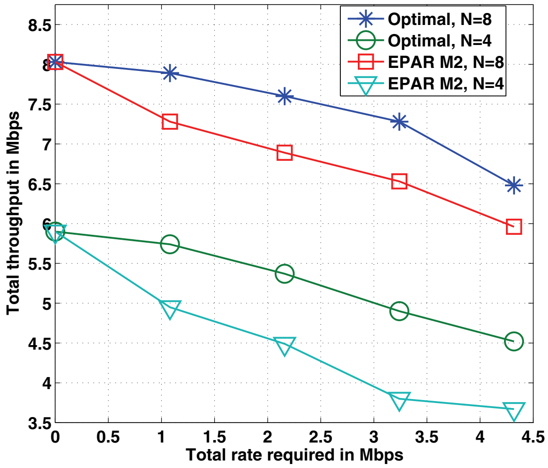

Fig. 10. Total system throughput for different rate requirements, $\delta = 0.4$.

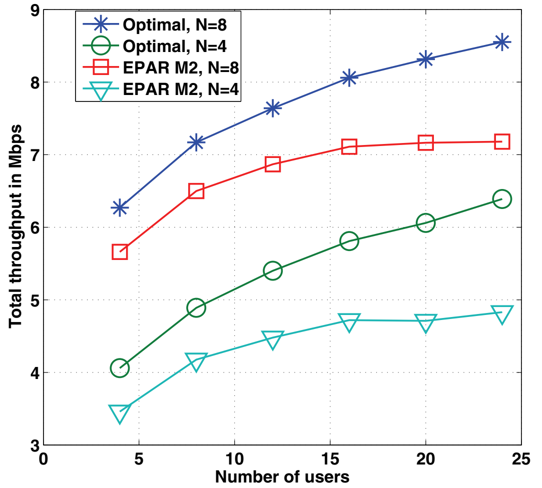

Fig. 9. Total system throughput with different number of users, $\delta = 0.4$.

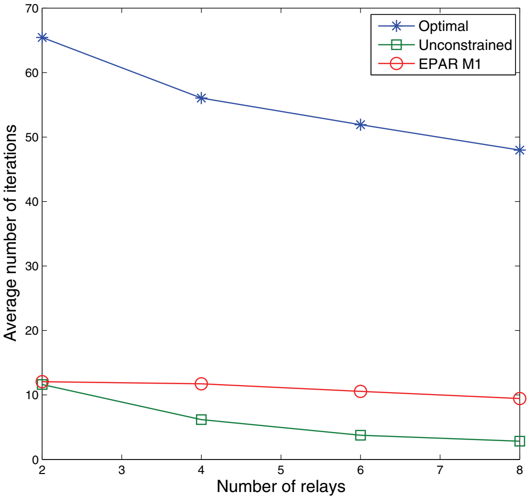

Fig. 11. Average number of iterations for converging all dual variables with $K = 16$, where 8 GBR users with equal rate requirements of 3 bits/sec/Hz.

The optimal scheme still provides moderate performance.

Fig. 6 shows the average rate obtained by each user for the optimal scheme, unconstrained scheme and EPAR M2 scheme. Since all suboptimal schemes have almost the same performance in case of GBR users, we only show the suboptimal scheme which has good overall throughput with lower computational complexity, i.e., EPAR M2 scheme. In this illustrative example, there are four GBR users with different minimum rate requirements. The minimum rate required for users 2 and 8 are 4 bits/sec/Hz and 2 bits/sec/Hz for users 4 and 6. The remaining users have no minimum rate requirements. It is observed that the minimum rate requirements are not fulfilled for users 2 and 4 when we use the unconstrained scheme. But both the optimal and EPAR M2 schemes satisfy the minimum rate requirement for all GBR users and support all other AMBR users. So, it can be concluded that our proposed optimal and suboptimal scheme with power refinement provide not only fairness and user satisfaction but also support heterogeneous demand as well. This will be more evident via satisfaction index (SI) in Fig. 7. $SI = 1$ means all users rate requirements are satisfied. The SI is much higher for the optimal scheme and EPAR M2 scheme compared with the unconstrained scheme. It is also observed that all users are satisfied in case of the optimal scheme and EPAR M2 scheme when the number of relays increases. But all users are not satisfied even for 8 relays in case of the unconstrained scheme due to the same reason as stated above. The EPAR M2 scheme exhibits slightly higher SI than the optimal scheme. Because the EPAR M2 scheme first allocates resources to the GBR users, and when all GBR users are satisfied, the remaining resources are then allocated to the AMBR users.

The average user throughput by varying the relay locations for different number of relays is presented in Fig. 8. It is noticed that for all cases, the average throughput using relays increases first for the lower value of δ until it reaches the maximum and then decreases for the larger value of δ. Because, when δ is small, relays are located close to the

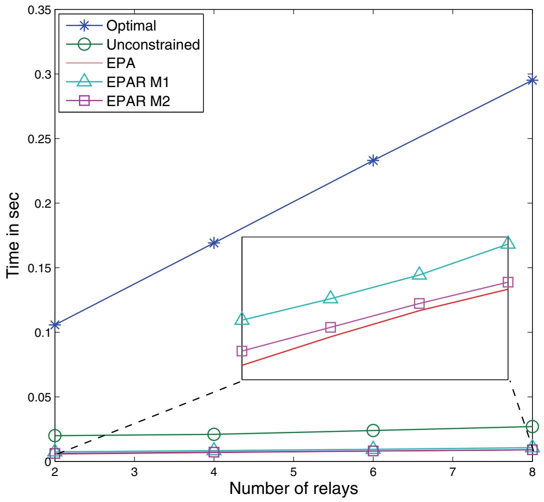

Fig. 12. Average running time for $K = 16$, where 8 GBR users with equal rate requirements of 3 bits/sec/Hz.

eNodeB. Hence, relays are not beneficial for the cell edge users due to low SNR of the SR links. Similarly, when relays are furthest from the eNodeB (i.e., high value of δ), the throughput reduces due to low SNR of the RD links.

Fig. 9 shows the total system throughput as a function of the number of users, with the number of relays as a parameter. The total throughput increases with the number of users or relays. This gives insight into the scalability of our schemes.

The total throughput as a function of the total rate required for all users is shown in Fig. 10. The figure reveals that the total throughput decreases with the increase of total required rate for both schemes. The reason is when the rate requirement of the GBR users continues to increase, the users and relays need to increase their rates by utilizing their maximum power and acquiring more subcarriers. When the total rate requirement is zero, i.e., there are no GBR users, both schemes behave like the unconstrained scheme and provide the same total throughput.

The complexity of the optimal scheme, EPAR M1 scheme and unconstrained scheme, mainly depends on the convergence of the dual variables. The complexity comparison for these schemes can be better illustrated by comparing the convergence of these schemes. However, the EPA scheme and EPAR M2 scheme do not require any update of dual variables. Therefore, we also present the running time comparison of all schemes.

The average number of iterations required to converge all the dual variables for the optimal scheme, EPAR M1 scheme and unconstrained scheme is shown in Fig. 11. The optimal scheme requires the highest number of iterations to converge all dual variables. This is due to the minimum rate constraints of each GBR user. In case of the optimal scheme, the number of dual variables is equal to the number of GBR users plus one for the total power constraint. On the other hand, the unconstrained scheme has only one dual variable, hence requires less number of iterations to converge. The EPAR M1 scheme has the same number of dual variables as the optimal scheme, however, it requires less number of iterations since it only reallocates power for a given subcarrier and relay assignment. It can also be seen that the number of iterations reduces with the increase of the number of relays. This is because when the number of relays increases, higher rates can be achieved by using more number of channels with good SNR, which reduces the number of iterations since the rate requirements for the GBR users can be easily obtained. However, the total running time of all these schemes increases with the increase of the number of relays, which is shown in Fig. 12.

The average running time of all schemes is shown in Fig. 12. The optimal scheme takes the largest amount of time to allocate resources since it requires large number of iterations to terminate all the dual variables. Between the suboptimal schemes, the EPAR M1 scheme takes the largest amount of time since it still needs to update the dual variables. The EPA scheme takes the least amount of time to allocate resources due to its simplicity. The running time of both the EPA and EPAR M2 schemes is very close while EPAR M2 scheme provides the highest throughput among all suboptimal schemes. The total time required to allocate resources for all schemes increases with the increase of the number of relays since the problem dimension and complexity cost increase.

VII. CONCLUSION

In this paper, relay selection and resource allocation in a multi-user cooperative OFDMA-based uplink system that simultaneously supports GBR and AMBR traffic have been investigated. A QoS aware optimal joint relay selection, power allocation and subcarrier assignment scheme under a total power constraint has been proposed. A joint optimization problem has been formulated for relay selection and resource allocation with the objective of maximizing the system throughput by satisfying the individual users' QoS requirements. By relaxing the integer constraints, the joint optimization problem has been transformed into a convex optimization problem, which is solved by means of a two level dual decomposition approach. The computational complexity has been finally reduced via the introduction of suboptimal schemes. Numerical results have demonstrated that our schemes support heterogeneous services while satisfy QoS requirements of each user. The polynomial complexity of the optimal scheme facilitates the implementation of this optimization at the base station. However, the suboptimal schemes can be implemented with significantly reduced computational complexity while sacrificing some system throughput. For the future work, we will investigate the performance of our schemes in the presence of imperfect CSI at the base station.

APPENDIX
PROOF OF PROPOSITION 1

According to the definition of subgradient [29], vector $c \in R^n$ is a subgradient of a given convex function $f : R^n \to R$ at the point $y \in R^n$ if $f(x) \geq f(y) + (x - y)^T c$, $\forall x \in R^n$.

Consider the objective function $g(\lambda, \mu)$ in (12) at two different points (λ, μ) and $(\lambda^{'}, \mu^{'})$, where $\lambda = (\lambda_1, \lambda_2, ..., \lambda_k, ...\lambda_K)$ and $\lambda^{'} = (\lambda_1, \lambda_2, ..., \lambda_k^{'}, ...\lambda_K), \forall k \in \kappa_1$. We have

$$g(\lambda,\mu) = \begin{cases} \max_{\rho,P_t} \quad L(\rho, P_t, \lambda, \mu) \\ s.t. \sum_{k=1}^{K} \sum_{n=0}^{N} \rho_{k,n}^{m} = 1, \forall m, \\ \quad 0 \geq \rho_{k,n}^{m} \leq 1, P_{t,k}^{m} \geq 0. \end{cases} \tag{28}$$

$$g(\lambda^{'},\mu^{'}) = \begin{cases} \max_{\rho,P_t} \quad L(\rho, P_t, \lambda^{'}, \mu^{'}) \\ s.t. \sum_{k=1}^{K} \sum_{n=0}^{N} \rho_{k,n}^{m} = 1, \forall m, \\ \quad 0 \geq \rho_{k,n}^{m} \leq 1, P_{t,k}^{m} \geq 0. \end{cases} \tag{29}$$

Substituting $\rho_{k,n}^{m}$ and $P_{t,k}^{m}$ with the optimal values, we have the subgradient of $g(\lambda,\mu)$ at λ_k

$$\begin{aligned} & [g(\lambda^{'},\mu^{'}) - g(\lambda,\mu)] \\ &= \max_{\rho,P_t} \quad L(\rho, P_t, \lambda^{'}, \mu^{'}) - \max_{\rho,P_t} \quad L(\rho, P_t, \lambda, \mu) \\ &\geq L(\rho^{\star}, P_t^{\star}, \lambda^{'}, \mu^{'}) - L(\rho^{\star}, P_t^{\star}, \lambda, \mu) \\ &= (\lambda_k^{'} - \lambda_k) \sum_{k \in \ \kappa 1} \left(\sum_{m=1}^{M} \sum_{n=0}^{N} \rho_{k,n}^{m^{\star}} R_{k,n}^{m^{\star}} - Q_k \right) \\ & \quad + (\mu^{'} - \mu)(P_T - \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{n=0}^{N} \rho_{k,n}^{m^{\star}} P_{t,k}^{m^{\star}}) \end{aligned} \tag{30}$$

The inequality in (30) holds because of the definition of dual function and Lagrange in (11) and (12), respectively. Thus, we have

$$\begin{aligned} g(\lambda^{'},\mu^{'}) \quad &\geq \quad g(\lambda,\mu) + (\lambda_k^{'} - \lambda_k) \sum_{k \in \ \kappa 1} \left(\sum_{m=1}^{M} \sum_{n=0}^{N} \rho_{k,n}^{m^{\star}} R_{k,n}^{m^{\star}} \right. \\ &- \quad \left. Q_k \right) + (\mu^{'} - \mu)(P_T - \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{n=0}^{N} \rho_{k,n}^{m^{\star}} P_{t,k}^{m^{\star}}). \end{aligned} \tag{31}$$

So, the subgradients of $g(\lambda,\mu)$ at the point λ_k are

$$\Delta\lambda_k = \sum_{m=1}^{M} \sum_{n=0}^{N} \rho_{k,n}^{m^{\star}} R_{k,n}^{m^{\star}} - Q_k, \forall k \in \kappa_1,$$

$$\Delta\mu = P_T - \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{n=0}^{N} \rho_{k,n}^{m^{\star}} P_{t,k}^{m^{\star}}.$$

REFERENCES

[1] M. S. Alam, J. W. Mark, and X. Shen, "Relay selection and resource allocation for multi-user cooperative LTE-A uplink," in *Proc. 2012 IEEE International Conf. Commun.*, pp. 5092–5096.

[2] S. Parkvall, A. Furuska, and E. Dahlman, "Evolution of LTE toward IMT-Advanced," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 84–91, Feb. 2011.

[3] 3GPP, TS-25.101 v6.4.0, "Technical specification group radio access network; User Equipment (UE) radio transmission and reception," pp. 1–272, Mar. 2004.

[4] 3GPP, TR-36.913 v8.0.1, "Requirements for further advancements for E-UTRA," pp. 1–17, Mar. 2009.

[5] IEEE Std 802.16m-2011, "IEEE standard for local and metropolitan area networks part 16: air interface for broadband wireless access systems amendment 3: advanced air interface," pp. 1–1112, June 2011.

[6] W. Zhuang and M. Ismail, "Cooperation in wireless communication networks," *IEEE Wireless Commun.*, vol. 19, no. 2, pp. 10–20, Apr. 2012.

[7] F. Hou, J. She, P.-H. Ho, and X. Shen, "A flexible resource allocation and scheduling framework for non-real-time polling service in IEEE 802.16 networks," *IEEE Trans. Wireless Commun*, vol. 8, no. 2, pp. 766–775, Feb. 2009.

[8] M. Mehrjoo, M. Awad, M. Dianati, and X. Shen, "Design of fair weights for heterogeneous traffic scheduling in multichannel wireless networks," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2892–2902, Oct. 2010.

[9] M. Awad, V. Mahinthan, M. Mehrjoo, X. Shen, and J. Mark, "A dual-decomposition-based resource allocation for OFDMA networks with imperfect CSI," *IEEE Trans. Veh. Technol.*, vol. 59, no. 5, pp. 2394–2403, June 2010.

[10] W. Dang, M. Tao, H. Mu, and J. Huang, "Subcarrier-pair based resource allocation for cooperative multi-relay OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1640–1649, May 2010.

[11] J. Cai, X. Shen, J. Mark, and A. Alfa, "Semi-distributed user relaying algorithm for amplify-and-forward wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1348–1357, Apr. 2008.

[12] S. Kadloor and R. Adve, "Relay selection and power allocation in cooperative cellular networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1676–1685, May 2010.

[13] M. Mehrjoo, S. Moazeni, and X. Shen, "Resource allocation in OFDMA networks based on interior point methods," *Wireless Commun. Mobile Comput.*, vol. 10, no. 11, pp. 1493–1508, 2010.

[14] J. Xu, X. Shen, J. Mark, and J. Cai, "Quasi-optimal channel assignment for real-time video in OFDM wireless systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1417–1427, Apr. 2008.

[15] A. T. Gamage, M. S. Alam, X. Shen, and J. Mark, "Joint relay, subcarrier and power allocation for OFDMA-based femtocell networks," in *Proc. 2013 IEEE Wireless Commun. Netw. Conf.*, pp. 851–856.

[16] L. You, P. Wu, M. Song, J. Song, and Y. Zhang, "Cross-layer optimisation for uplink transmission in OFDMA cellular networks with fixed relays," *European Trans. Telecommun.*, vol. 22, pp. 296–314, Oct. 2011.

[17] K. Ma, Z. Liu, and X. Guan, "Joint relay selection and power allocation for cooperative cellular networks," *Wireless Personal Commun.*, pp. 1–17, Dec., 2010.

[18] K. Sung, Y. P. Hong, and C. Chao, "Resource allocation and partner selection for cooperative multicarrier systems," *IEEE Trans. Veh. Technol.*, vol. 60, no. 7, pp. 3228–3240, Sept. 2011.

[19] V. Mahinthan, L. Cai, J. Mark, and X. Shen, "Partner selection based on optimal power allocation in cooperative-diversity systems," *IEEE Trans. Veh. Technol.*, vol. 57, no. 1, pp. 511–520, Jan. 2008.

[20] P. Thulasiraman and X. Shen, "Interference aware resource allocation for hybrid hierarchical wireless networks," *Computer Netw.*, vol. 54, pp. 2271–2280, Sept. 2010.

[21] C.-N. Hsu, H.-J. Su, and P.-H. Lin, "Joint subcarrier pairing and power allocation for OFDM transmission with decode-and-forward relaying," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 399–414, Jan. 2011.

[22] D. Ng and R. Schober, "Cross-layer scheduling for OFDMA amplify-and-forward relay networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 3, pp. 1443–1458, Mar. 2010.

[23] W. C. Pao and Y. F. Chen, "Reduced complexity subcarrier allocation schemes for DFT-precoded OFDMA uplink systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 9, pp. 2701–2706, Sept. 2010.

[24] X. Zhang, S. Chen, and W. Wang, "Multiuser radio resource allocation for multiservice transmission in OFDMA-based cooperative relay networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2009, pp. 1–13, 2009.

[25] Y. Li, W. Wang, J. Kong, W. Hong, X. Zhang, and M. Peng, "Power allocation and subcarrier pairing in OFDM-based relaying networks," in *Proc. 2008 IEEE International Conf. Commun.*, pp. 2602–2606.

[26] M. Hajiaghayi, M. Dong, and B. Liang, "Optimal channel assignment and power allocation for dual-hop multi-channel multi-user relaying," in *Proc. 2011 IEEE International Conf. Comput. Commun.*, pp. 76–80.

[27] L. Vandendorpe, R. Duran, J. Louveaux, and A. Zaidi, "Power allocation for OFDM transmission with DF relaying," in *Proc. 2008 IEEE International Conf. Commun.*, pp. 3795–3800.

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

[29] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar , *Convex Analysis and Optimization*. Athena Scientic, 2003.

[30] D. Zhang, Y. Wang, and J. Lu, "QoS aware relay selection and subcarrier allocation in cooperative OFDMA systems," *IEEE Commun. Lett.*, vol. 14, no. 4, pp. 294–296, Apr. 2010.

Md Shamsul Alam (S'10) received the B.Sc. in Electrical and Electronic Engineering from Rajshahi University of Engineering and Technology (RUET), Bangladesh and M.Sc. in Electrical and Computer Engineering from University of Calgary,Canada, in 2000 and 2009, respectively. He is a research assistant and currently working towards his Ph.D. degree at the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include resource allocation and quality-of-service provisioning in cooperative cellular networks, and integrated wireless networks. He served as a technical reviewer for several conferences and journals (IEEE JSAC and *IET Communications*). He was an Assistant Professor in the Department of Computer Science and Engineering at RUET, Bangladesh from 2004 to 2006. He was the recipient of TRLabs Canada Telecommunication Research Scholarship and Institute Gold Medel from RUET, Bangladesh.

Jon W. Mark (M'62-SM'80-F'88-LF'03) received the B.A.Sc. degree from the University of Toronto in 1962, and the M.Eng. and Ph.D. degrees from McMaster University, Canada in 1968 and 1970, respectively, all in electrical engineering. From 1962 to 1970, he was an engineer and then senior engineer at Westinghouse Canada Ltd. where he conducted research in advanced sonar signal processing. From 1968 to 1970, he was on leave from Westinghouse to pursue Ph.D. studies at McMaster University under the auspices of an NRC PIER Fellowship. He joined the Department of Electrical Engineering at the University of Waterloo in September 1970, and was promoted to the rank of full Professor in July 1978. He served as Department Chairman from July 1984 to June 1990. During this period, the department introduced the computer engineering degree program and changed the name to Electrical and Computer Engineering. He was on sabbatical leaves at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, as a Visiting Research Scientist (1976-77), at Bell Labs, Murray Hill, NJ, as a Resident Consultant (1983-84), at the Université Pierre et Marie Curie, Paris, France, as an Invited Professor (1990-91), and at the National University of Singapore as a Visiting Professor (1994-95). He established the Centre for Wireless Communications (CWC) at the University of Waterloo in 1996 with a $1 million donation from Ericsson Canada as seed money. He is currently a Distinguished Professor Emeritus and the founding Director of CWC at the University of Waterloo. He received the 2000 Canadian Award in Telecommunications Research for significant research contributions, scholarship and leadership in the fields of computer communication networks and wireless communications and the 2000 Award of Merit by the Educational Foundation of the Association of Chinese Canadian Professionals for Significant Contributions in Telecommunications Research. His current research interests are in wireless communications and wireless/wireline interworking, with a focus on cooperative networking, power control and resource allocation, etc. He is a co-author of *Wireless Communications and Networking* (Prentice-Hall, 2003), *Multimedia Services in Wireless Internet* (John Wiley & Sons, 2009), and *Wireless Broadband Networks* (John Wiley & Sons, 2009).

Dr. Mark is a Life Fellow of the IEEE and has served as a member of a number of editorial boards, including editorships in IEEE TRANSACTIONS ON COMMUNICATIONS, Wireless Networks, and Telecommunication Systems, a member of the Inter-Society Steering Committee of the IEEE/ACM TRANSACTIONS ON NETWORKING during the period 1992-2003 (as the SC Chair during 1999-2000), and a member of the IEEE Communications Society Awards Committee during 1995-1998. He is currently a member of the Advisory Board, John Wiley Series on Advanced Texts in Communications and Networking.

Xuemin (Sherman) Shen (IEEE M'97-SM'02-F'09) received the B.Sc. (1982) degree from Dalian Maritime University (China) and the M.Sc. (1987) and Ph.D. degrees (1990) from Rutgers University, New Jersey (USA), all in electrical engineering. He is a Professor and University Research Chair, Department of Electrical and Computer Engineering, University of Waterloo, Canada. He was the Associate Chair for Graduate Studies from 2004 to 2008. Dr. Shen's research focuses on resource management in interconnected wireless/wired networks, wireless network security, wireless body area networks, vehicular ad hoc and sensor networks. He is a co-author/editor of six books, and has published more than 600 papers and book chapters in wireless communications and networks, control and filtering. Dr. Shen served as the Technical Program Committee Chair for IEEE VTC'10 Fall, the Symposia Chair for IEEE ICC'10, the Tutorial Chair for IEEE VTC'11 Spring and IEEE ICC'08, the Technical Program Committee Chair for IEEE Globecom'07, the General Co-Chair for Chinacom'07 and QShine'06, the Chair for IEEE Communications Society Technical Committee on Wireless Communications, and P2P Communications and Networking. He also serves/served as the Editor-in-Chief for *IEEE Network*, *Peer-to-Peer Networking and Application*, and *IET Communications*; a Founding Area Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS; an Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *Computer Networks*, and *ACM/Wireless Networks*, etc.; and the Guest Editor for IEEE JSAC, *IEEE Wireless Communications*, *IEEE Communications Magazine*, and *ACM Mobile Networks and Applications*, etc. Dr. Shen received the Excellent Graduate Supervision Award in 2006, and the Outstanding Performance Award in 2004, 2007 and 2010 from the University of Waterloo, the Premier's Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada, and the Distinguished Performance Award in 2002 and 2007 from the Faculty of Engineering, University of Waterloo. Dr. Shen is a registered Professional Engineer of Ontario, Canada, an IEEE Fellow, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, and a Distinguished Lecturer of IEEE Vehicular Technology Society and Communications Society. Dr. Shen has been a guest professor of Tsinghua University, Shanghai Jiao Tong University, Zhejiang University, Beijing Jiao Tong University, Northeast University, etc.