Semi-Distributed User Relaying Algorithm for Amplify-and-Forward Wireless Relay Networks

Jun Cai, *Member, IEEE,* Xuemin (Sherman) Shen, *Senior Member, IEEE,* Jon W. Mark, *Life Fellow, IEEE,* and Attahiru S. Alfa, *Member, IEEE*

***Abstract*— In this paper, designing an effective user relaying algorithm, in terms of relay node selection and power allocation, is discussed for amplify-and-forward wireless relay networks. The objective is to simplify the application of user relaying in practical wireless communication networks so that the system capacity can be improved with low computational complexity and system overhead. Beginning with the derivation of a tight threshold-based sufficient condition on the feasibility of a relay node, i.e., ensuring that user relaying via the node can achieve a larger channel capacity than direct transmission, a semi-distributed user relaying algorithm is proposed. In the proposed algorithm, each relay node can make decision on its feasibility individually, and the ultimate decision on the relay node selection among multiple feasible ones is made in a centralized manner. Since there is no need on exchanging channel state information among different network nodes, the proposed algorithm is simple for implementation and suitable for practical applications, which have stringent constraints on system overhead. By comparing with the centralized user relaying algorithm, which requires global channel state information of the whole network, the proposed semi-distributed algorithm can provide comparable system capacity, but has significantly reduced computational complexity.**

***Index Terms*— User relaying, amplify-and-forward, power allocation, scheduling, wireless relay networks.**

I. INTRODUCTION

WITH the development of advanced wireless communication technologies, wireless networks have been widely accepted as an inexpensive solution to provide last-mile access to the Internet [1]. The concept of wireless backbone networks has appeared in the current standards, such as IEEE 802.16 or WiMAX (worldwide interoperability for microwave access). Different from wireline networks, broadcast is an inherent property of wireless transmission, i.e., the information transmitted from a source node can be received by not only the destination node, but also the neighboring nodes surrounding the source. In traditional wireless networks, such signals received by the neighboring nodes are treated as interference and many communication techniques (such as equalization) have been developed to alleviate their impact. However, such signals actually contain the same information as that received by the destination node. Therefore, if they can be properly forwarded to the destination, the reception performance may be improved. This motivates the application of a new technology, called user relaying.

User relaying is an emerging technology to achieve virtual spatial diversity in wireless communication networks [2]. The basic idea is that some nodes in the network overhear the information transmitted from the source node and relay what is received to the destination. Since the same information is obtained at the destination node from different spatially located nodes, a kind of spatial diversity, called cooperative diversity, is achieved. Compared with multi-antenna technique, a traditional way to achieve spatial diversity, user relaying can provide spatial diversity gain without requiring the establishment of multiple antennas on the source node, which ordinarily has very limited space. Recently, cooperative diversity has been considered as a promising technique for ad hoc, sensor, and mesh networks, and has been expected to be involved in the standards, such as IEEE 802.11, IEEE 802.15.4a, and IEEE 802.16 [3]–[5].

User relaying could be implemented using *amplify-and-forward* [6] or *decode-and-forward* approach [6], [7]. In amplify-and-forward, the relay node simply amplifies and forwards what is received from the source node, while in decode-and-forward, the relay node will first verify the correctness of the message from the source node by decoding, and then forwards the re-encoded signal to the destination. Each of the two approaches has its advantages and disadvantages for communications in wireless environments. In this paper, the discussion will focus on amplify-and-forward relay networks only since it has advantages in simple implementation and low computation load on relay nodes, .

The effectiveness of the user relaying on network performance improvement has been demonstrated in [6], [8]–[11]. However, analytical results also indicate that such performance improvement heavily depends on selecting suitable relay nodes and allocating proper power among the source and the relay nodes [6]. Otherwise, applying user relaying may not achieve performance improvement, or may even worsen the performance, compared to direct transmission without relaying. Therefore, designing effective relay node selection and power allocation algorithms becomes critical in wireless relay networks. However, the design is challenging in practical scenarios, which have limitations on the availability of channel state information and the system complexity. Recent research work in the literature can be classified into two groups. One group focuses on designing distributed space-time coding algorithms [12]. While fully distributed relay node selection can

Manuscript received November 7, 2006; revised February 13, 2007; accepted February 27, 2007. The associate editor coordinating the review of this paper and approving it for publication was D. Wu. The work of J. Cai and A. S. Alfa has been supported in part by a Natural Sciences and Engineering Research Council (NSERC) of Canada IRC Grant. The work of X. Shen and J. W. Mark has been supported by NSERC Discovery Grants under grant nos. RGPIN203560 and RGPIN7779.

J. Cai and A. S. Alfa are with the Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, Manitoba, Canada, R3T 3N8 (e-mail: {jcai, alfa}@ee.umanitoba.ca).

X. Shen and J. W. Mark are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1 (e-mail: {xshen, jwmark}@bbcr.uwaterloo.ca).

Digital Object Identifier 10.1109/TWC.2008.060909.

be achieved, such coding schemes require complete channel state information between any pair of nodes, which is ordinarily infeasible in practical networks. In addition, encoding and decoding may result in high computational overhead. The other group designs explicit relay node selection and power allocation algorithms, such as [6], [13]–[17]. Relay node selection, referred to as partner matching, is studied in [18]. Both centralized [6], [13], [14] and distributed algorithms [15]–[17] have been proposed. The centralized algorithms focus on solving complicated optimization problems by assuming complete channel state information, which makes the centralized algorithms impractical to implement. The distributed algorithms are based on opportunistic cooperation, which may induce high system overhead due to contention. In addition, few works [19] consider the application of user relaying from the whole network point of view. Therefore, it is still an important and open topic to design a practically implementable user relaying algorithm for wireless communication networks.

In this paper, designing effective user relaying algorithms by jointly optimizing relay node selection and power allocation is proposed for wireless relay networks. A simple relay network, which consists of one source and destination pair, is first studied. A joint optimization problem, which integrates relay node selection and power allocation to maximize network capacity in terms of achievable mutual information, is formulated and the optimal solutions are obtained by completely searching all available relay nodes. In order to reduce the computational complexity, the original joint optimization problem is decoupled into two subproblems, which include best relay node searching with uniform power distribution between the source node and the relay node, and the optimal power allocation based on the given relay node. Under uniform power allocation, both sufficient and necessary conditions are derived for determining the feasibility of the relay nodes, i.e., applying user relaying through these relay nodes will provide better capacity than direct transmission. Two thresholds are provided to the sufficient and necessary conditions, which significantly simplifies the search for the best relay node. Then, a closed-form optimal power allocation scheme is developed given the relay node selection. The obtained results are then applied to a more general relay network with multiple source and destination pairs, and a semi-distributed user relaying algorithm is proposed. By "semi-distributed" it is meant that each relay node can individually decide on its feasibility, while the final decision on the relay node selection is still carried out in a centralized manner. Since there is no requirement on exchanging channel state information, the proposed algorithm is simple for implementation and suitable for practical wireless communication networks, which have stringent constraints on system overhead. By comparing with the optimal case, where complete channel state information is available, the proposed semi-distributed algorithm can provide comparable capacity performance with significantly reduced computational complexity, measured in terms of average search time.

The remainder of this paper is organized as follows. In Section II, the system model of the wireless relay network under consideration is defined and the system capacity in terms of achievable maximum average mutual information is formulated. User relaying algorithms for the network with a single source and destination pair are discussed in Section III, where sufficient and necessary conditions for relay node selection and closed-form power allocation are derived. Section IV presents the proposed semi-distributed user relaying algorithm for the network with multiple source and destination pairs. In Section V, the performance of the proposed user relaying algorithms is evaluated by computer simulation. Conclusions of this work are given in Section VI.

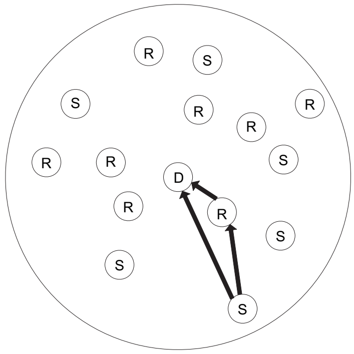

Fig. 1. System model of the wireless relay network.

II. System Model

Fig. 1 shows the system model of the wireless relay network under consideration, which consists of N source (S) nodes, N' available relay (R) nodes, and a single destination (D) node. One interpretation of such model is the IEEE 802.16 point-to-multiple-point (PMP) network, where the source nodes represent the end users with packets waiting for transmission, and the destination node represents the access point (AP). Nonetheless, most results derived in this paper are generally applicable to networks with multiple destination nodes. In this study, the inter-user interference is omitted by allocating orthogonal channels for different source nodes, e.g., using orthogonal frequency division multiple access (OFDMA). How to combat the inter-user interference is beyond the scope of this paper and will be considered in the future. For each source node, except for the direct connection to the destination node, one of the available relay nodes can be chosen to construct a relay path. By appropriately selecting a suitable relay node and distributing transmission power, effective cooperative diversity gain can be achieved to improve the system capacity. The number of relay nodes selected for each source node is limited to no more than 1. Although employing multiple relay nodes at the same time may further improve the cooperative diversity gain, the resultant lower spectrum efficiency and high computational complexity may lead to beneficial performance-complexity tradeoff [6]. Each relay node works under the amplify-and-forward mode, i.e., the relay node simply amplifies the received signal from the source node and then forwards it to the destination.

The transmission in the time domain is on a frame-by-frame basis. Each frame consists of two consecutive time slots. The first time slot is used for the transmissions of the source nodes and the second one is used for either relay nodes or source nodes depending on whether user relaying is applied. Such transmission pattern reflects the half-duplex requirement of the practical equipment, which ordinarily cannot transmit and receive at the same time.

Each pair of nodes, i and j, experiences a flat fading channel with channel gain α_{ij} [6]. The channel gain may integrate the effects from both propagation path loss and fading, and is slowly time-variant so that it is approximately unchanged during each frame interval. Since this paper focuses on designing relay node selection and power allocation algorithms, we omit the discussion on channel estimation and assume the availability of perfect channel state information. For channel estimation methods, the reader is referred to works available in the literature, e.g., [20], for channel estimation in OFDM systems.

Consider a transmission consisting of a source node s, a relay node i and a destination node d. Let a frame begin at time slot n. Then, at the end of the slot, the received signals at relay node i and destination node d are respectively given by

$$y_{si}[n] = \alpha_{si}x_s[n] + Z_i[n], \quad (1)$$
$$y_{sd}[n] = \alpha_{sd}x_s[n] + Z_d[n] \quad (2)$$

where α_{si} and α_{sd} denote the channel gains from the source node to the relay node and from the source node to the destination node, respectively. $x_s[n]$ denotes the information signal transmitted from the source node at time slot n. $Z_i[n]$ and $Z_d[n]$ denote the background noise at relay node i and destination node d, respectively, which are independent identically distributed (i.i.d.) complex Gaussian random variables with a common variance σ_n^2.

At the end of time slot $n+1$, the received signal at destination node d from relay node i is given by

$$\begin{aligned} y_{id}[n+1] &= \alpha_{id}\lambda_i y_{si}[n] + Z_d[n+1] \\ &= \alpha_{id}\lambda_i\alpha_{si}x_s[n] + \alpha_{id}\lambda_i Z_i[n] + Z_d[n+1] \end{aligned} \quad (3)$$

where α_{id} is the channel gain between the relay node and the destination node. λ_i is an amplification factor, which is used to guarantee the transmission power of the relay node and satisfies

$$\begin{aligned} &\lambda_i^2(|\alpha_{si}|^2 P_s + \sigma_n^2) = P_i \\ \Rightarrow \quad &\lambda_i^2 = \frac{P_i}{|\alpha_{si}|^2 P_s + \sigma_n^2} \end{aligned} \quad (4)$$

where P_s and P_i denote the transmission power of the source node and the relay node, respectively.

Equations (1), (2), and (3) can be combined into a more compact form as

$$\mathbf{Y} = \mathbf{A}x_s[n] + \mathbf{BZ} \quad (5)$$

where

$$\begin{aligned} \mathbf{Y} &= [y_{sd}[n], y_{id}[n+1]]^T \\ \mathbf{A} &= [\alpha_{sd}, \alpha_{id}\lambda_i\alpha_{si}]^T \\ \mathbf{B} &= \begin{bmatrix} 0 & 1 & 0 \\ \alpha_{id}\lambda_i & 0 & 1 \end{bmatrix} \\ \mathbf{Z} &= [Z_i[n], Z_d[n], Z_d[n+1]]^T \end{aligned}$$

and the superscript "T" denotes matrix transpose.

Channel capacity in bits per unit bandwidth is defined as the average mutual information maximized over all choices of the marginal probabilities, while the actual capacity is given by the product of maximum average mutual information and the system bandwidth. If the entire system capacity can be utilized to delivery messages, then throughput is identical to system capacity.

If α_{si}, α_{id} and α_{sd} are known, the channel combining both the direct path and the relay path can be modelled as an equivalent one-input, two-output complex Gaussian noise channel, which has the maximum average mutual information I_{AF}^i given by [6]

$$I_{AF}^i = \frac{1}{2}\log\det(\mathbf{I} + (P_s\mathbf{AA}^H)(\mathbf{B}E[\mathbf{ZZ}^H]\mathbf{B}^H)^{-1}) \quad (6)$$

where $E[\cdot]$ denotes the mean of the random variable, $\det(\cdot)$ outputs the determinant of the matrix, $\mathbf{I}$ is an identity matrix, and the superscript "H" represents the complex conjugate transposition. Let $\beta_{si} = |\alpha_{si}|^2$, $\beta_{sd} = |\alpha_{sd}|^2$, and $\beta_{id} = |\alpha_{id}|^2$. We have

$$I_{AF}^i = \frac{1}{2}\log(1 + \frac{P_s}{\sigma_n^2}(\beta_{sd} + \frac{\beta_{si}\beta_{id}P_i}{\beta_{si}P_s + \beta_{id}P_i + \sigma_n^2})). \quad (7)$$

Note that λ_i in (7) has been replaced by (4).

If no relaying is applied and both time slots in one frame contribute to the source node's transmissions, the achievable average mutual information for the direct transmission can be calculated as

$$\begin{aligned} I_D &= 2 \times \frac{1}{2}\log(1 + \frac{\beta_{sd}P_s}{\sigma_n^2}) \\ &= \frac{1}{2}\log(1 + \frac{\beta_{sd}P_s}{\sigma_n^2})^2. \end{aligned} \quad (8)$$

Combining (7) and (8), the maximum average mutual information (or capacity) for a pair of cooperating users is equal to

$$C = \max\{I_D, \max_i(I_{AF}^i)\}. \quad (9)$$

Obviously, the achievable maximum average mutual information depends on the transmission power of the source node (P_s), the transmission power of the relay node (P_i), and the channel gains (β_{id} and β_{sd}). Therefore, in order to achieve maximum capacity, all parameters should be appropriately considered.

III. Relay Network with Single Source Node

In this section, we consider a relay network, which consists of one single source node.

A. Optimization Problem Formulation

Our objective is to maximize the capacity of the relay network by designing suitable values of P_s and P_i, and selecting a relay node with appropriate channel gains β_{si} and β_{id}. The constraint is that the total transmission power, P_s+P_i, is upper bounded by P_{max}. Such constraint is the common requirement in practical systems where the transmission power should be properly distributed to lengthen the network lifetime by avoiding power exhaustion on only some of the nodes. According to (7) and (8), the optimization problem should address the following questions:

1) Does the user relaying outperform the direct transmission?
2) Which relay node should be selected to achieve maximum capacity?

3) How is the total transmission power distributed between the source node and the relay node?

By taking all these aspects into account, the optimization problem in terms of relay node selection and power allocation can be formulated as

$$\begin{cases} \max\{I_D, \max_{\epsilon_i,P_s,P_i}\{\sum_{i\in\mathbf{\Gamma}_a}\epsilon_i I_{AF}^i\}\} \\ \text{subject to} \\ P_s + \sum_{i\in\mathbf{\Gamma}_a}\epsilon_i P_i \le P_{max}, & \text{(condition I)} \\ \sum_{i\in\mathbf{\Gamma}_a}\epsilon_i \le 1, & \text{(condition II)} \\ P_s, P_i \ge 0, \quad \forall i \in \mathbf{\Gamma}_a \end{cases} \tag{10}$$

where $\mathbf{\Gamma}_a$ denotes the set of all available relay nodes. ϵ_i is an indicator function of relay node selection, which takes a value of either 0 or 1. $\epsilon_i = 1$ means that relay node i is selected for the given source-destination pair and $\epsilon_i = 0$ otherwise. In (10), condition I is the constraint on the total transmission power and condition II means that at most one relay node can be selected. If $\sum_i \epsilon_i = 0$, both time slots are used for the transmissions of the source node, i.e., direct transmission.

Since the size of $\mathbf{\Gamma}_a$ is ordinarily limited in practice, i.e., $|\mathbf{\Gamma}_a| = N' < \infty$, the optimal solution of (10) can be obtained by completely searching all available relay nodes and then choosing the one which can provide the maximum capacity with optimal power allocation. We call such an algorithm the *complete searching algorithm*. Obviously, the complexity of the complete searching algorithm equals N' times of the computational efforts required to solve the optimal power allocation problem for each relay node. Thus, if the number of available relay nodes is large, the computational complexity of the complete searching algorithm may introduce heavy burden on the destination node's operation, especially for a battery-powered node. The situation becomes worse in a network consisting of N ($N >> 1$) source nodes. In other words, we need to solve $N' \times N$ optimal power allocation problems before obtaining the optimal solution for (10). In order to reduce the computational complexity, in the following, we decouple the original joint optimization problem in (10) into two subproblems. One is to search the best relay node based on a uniform power distribution between the source and the relay node, and the other is to optimize power distribution given that one relay node has been selected. Simulation results in Section V show that such decoupling does not degrade much the capacity performance compared to the complete searching algorithm but provides significant deduction on the computational complexity.

B. Relay Node Selection

In this subsection, the conditions for selecting a feasible relay node are derived under the assumption that the transmission power is uniformly distributed between the source node and the relay node. A relay node is feasible if user relaying via it can provide better capacity performance than direct transmission. The assumption of uniform power distribution is reasonable since uniform power distribution is optimal in practical systems if power allocation is not applicable or channel state information is unknown.

Since the average mutual information I_{AF}^i is an increasing function of both P_s and P_i as shown in (7), the maximum capacity can be achieved when equality is held in condition II of (10). Therefore, for uniform power distribution,

$$P_s = P_i = P_{max}/2 \triangleq P. \tag{11}$$

Substituting (11) into (7) and (8), we have

$$I_{AF}^i = \frac{1}{2}\log(1 + SNR(\beta_{sd} + \frac{\beta_{si}\beta_{id}SNR}{(\beta_{si}+\beta_{id})SNR+1})) \tag{12}$$

$$I_D = \frac{1}{2}\log(1+\beta_{sd}SNR)^2 \tag{13}$$

where $SNR = P/\sigma_n^2$ is the signal-to-noise ratio at the transmitting node.

If relay node i is feasible, then

$$\begin{aligned} &I_{AF}^i > I_D \\ \Rightarrow &1 + SNR(\beta_{sd} + \frac{\beta_{si}\beta_{id}SNR}{(\beta_{si}+\beta_{id})SNR+1}) > (1+\beta_{sd}SNR)^2 \\ \Rightarrow &\frac{\beta_{si}\beta_{id}SNR}{(\beta_{si}+\beta_{id})SNR+1} > \beta_{sd}(1+\beta_{sd}SNR). \end{aligned} \tag{14}$$

For the given source and destination pair, the right-hand side of the inequality is a constant. Let

$$C_n = \beta_{sd}(1+\beta_{sd}SNR). \tag{15}$$

Then, (14) becomes

$$\beta_{si}\beta_{id}SNR/C_n - (\beta_{si}+\beta_{id})SNR - 1 > 0. \tag{16}$$

Equation (16) can also be written as

$$\beta_{si}(\beta_{id}SNR/C_n - SNR) - (\beta_{id}SNR + 1) > 0. \tag{17}$$

Since all β_{si}, β_{id}, and SNR are larger than 0, the necessary condition that the inequality holds is that

$$\begin{aligned} &\beta_{id}SNR/C_n - SNR > 0 \\ \Rightarrow &\beta_{id} > C_n. \end{aligned} \tag{18}$$

Similarly, we have

$$\beta_{si} > C_n. \tag{19}$$

Define $\mathbf{\Gamma}_f$ as the set of feasible relay nodes. Combining (18) and (19) leads to the following Lamma.

Lamma 1

The necessary condition that a relay node i belongs to the feasible set is

$$\beta_{si} > C_n \text{ and } \beta_{id} > C_n, \quad \forall i \in \mathbf{\Gamma}_f.$$

From (16) and **Lamma 1**, some important observations can be obtained as follows.

1) β_{si} and β_{id} play the same role for better capacity. Therefore, it is expected that the relay node, which has one channel gain dominating the other (i.e., $\beta_{si} >> \beta_{id}$ or $\beta_{si} << \beta_{id}$), may not be the optimal solution for maximizing capacity even if it were feasible.
2) Only the relay nodes which satisfy the necessary condition in **Lamma 1** can be used to achieve better performance in terms of capacity over direct transmission. It

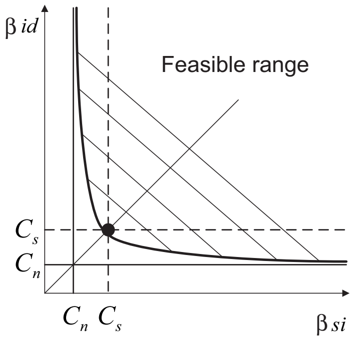

Fig. 2. Range of channel gains for the feasibility of relay nodes.

matches the general observation in the literature that the relay node should be carefully selected for performance improvement.

3) The necessary condition significantly reduces the searching space and, in turn, the searching time for the best relay node. As a result, the computational complexity can be effectively reduced.
4) The threshold C_n is an increasing function of β_{sd}. Therefore, if the channel condition between the source node and the destination node is good enough, user relaying cannot provide performance improvement. This observation matches the conclusions in [6]. Moreover, the threshold C_n is also an increasing function of SNR. Therefore, if the system has no power limitation, applying user relaying cannot offer capacity gain.
5) If the channel gain is dominated by the propagation path loss, the necessary condition is equivalent to limiting the relay nodes to locate within two circles, with the same radius C_n and centered at the source node and the destination node, respectively. In this case, the feasible relay nodes will locate in the intersection of the two circles.

Fig. 2 shows the feasible range of β_{si} and β_{id} based on (16) and the definition of C_n in (15). Obviously, since **Lamma 1** only defines the necessary condition for the feasibility of relay nodes, there exist relay nodes which satisfy the necessary condition but are not feasible. By observing Fig. 2 and considering the equal effect of the two channel gains, we define a new threshold C_s to further reduce the searching time. C_s can be obtained by solving the following equation,

$$\begin{cases} \beta_{si}\beta_{id}SNR/C_s - (\beta_{si}+\beta_{id})SNR - 1 > 0 \\ \\ \beta_{si} = \beta_{id}. \end{cases} \quad (20)$$

It can be shown that

$$C_s = \left(1 + \sqrt{1 + \frac{1}{SNR \cdot C_n}}\right) C_n. \quad (21)$$

By applying C_s as the threshold, the relay nodes satisfying $\beta_{si} > C_s$ and $\beta_{id} > C_s$ must achieve a better capacity than direct transmission or be feasible. Only the relay nodes, which are logically very close to the source node ($\beta_{si} >> \beta_{id}$) or the destination node ($\beta_{si} << \beta_{id}$), are omitted for consideration. By considering the equal effects of β_{si} and β_{id} on the capacity, we can expect that omitting such nodes will not significantly affect the ultimate performance. A sufficient condition for relay node feasibility is summarized as follows.

Lamma 2
The sufficient condition that any relay node i belongs to the feasible set is

$$\beta_{si} > C_s \text{ and } \beta_{id} > C_s, \quad \forall i \in \mathbf{\Gamma}_f. \quad (22)$$

Define $\mathbf{\Gamma}_n$ and $\mathbf{\Gamma}_s$ as the sets of relay nodes satisfying the necessary condition and the sufficient condition, respectively. Then,

$$\mathbf{\Gamma}_s \subset \mathbf{\Gamma}_f \subset \mathbf{\Gamma}_n. \quad (23)$$

We summarize the searching algorithm based on the sufficient condition for the best relay node to achieve the maximum capacity as follows.

- If $\mathbf{\Gamma}_s$ is not empty, then the best relay node is the one such that
$$i = \arg\max_{j\in\mathbf{\Gamma}_s} \left\{ \frac{\beta_{sj}\beta_{jd}SNR}{(\beta_{sj}+\beta_{jd})SNR+1} \right\}; \quad (24)$$
- Otherwise, the best choice is to apply direct transmission for both time slots in the frame.

C. Power Allocation

After selecting the best relay node, power allocation can be applied to further improve the capacity performance. Let

$$f(P_s, P_i) \triangleq \frac{P_s}{\sigma_n^2}(\beta_{sd} + \frac{\beta_{si}\beta_{id}P_i}{\beta_{si}P_s + \beta_{id}P_i + \sigma_n^2}). \quad (25)$$

According to (10), the optimal power allocation can be formulated as

$$\begin{cases} \max_{P_s,P_i} f(P_s,P_i) \\ \\ \text{subject to} \\ \\ P_s + P_i = P_{max} \\ \\ P_s, P_i > 0. \end{cases} \quad (26)$$

If optimal solutions exist, they can be obtained by solving the equation

$$\frac{df(P_s, P_{max} - P_s)}{dP_s} = 0. \quad (27)$$

A long tedious computation shows that there exist two closed-form solutions for the optimization problem in (26). The two solutions correspond to allocating more power to the source node or to the relay node. In practice, one of them can be selected by considering the remaining battery lifetime of the source or the relay node.

IV. Relay Networks with Multiple Source Nodes

In this section, we apply the results obtained in the previous section to design a semi-distributed user relaying algorithm for a more general relay network, which consists of multiple source nodes. For description simplicity, we omit the consideration of power allocation and assume uniform power distribution between any pair of source and relay nodes.

All the nodes in the network are partitioned into two groups. One group consists of source nodes, which have packets waiting for transmission to the destination, and the other includes nodes with relay function only. Let N and N'

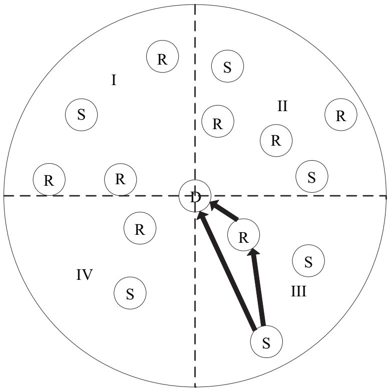

Fig. 3. Clustering method for simplifying exclusive searching algorithm.

represent the numbers of source nodes and relay nodes in the network, respectively. Note that there is no limitation on the value of N or N', i.e., N may be larger than, equal to, or less than N'. How to allocate N' relay nodes to N source nodes is the major problem we will discuss in this Section.

If complete channel state information is available at the destination node, we can formulate an optimization problem to maximize the network capacity as follows.

$$\begin{cases} \sum_{k=1}^{N} \max\{\max_{A_{k,i}}\{A_{k,i} \cdot I^{i}_{AF,k}\}, I_{D,k}\} \\ \text{subject to} \\ \sum_{k=1}^{N} A_{k,i} \leq 1, \quad \text{(condition I)} \\ \sum_{i=1}^{N'} A_{k,i} \leq 1, \quad \text{(condition II)} \\ A_{k,i} \in \{0,1\} \end{cases} \tag{28}$$

where the subscripts k and i denote the index of the source node and the relay node, respectively. $A_{k,i}$ is an indicator function. $A_{k,i} = 1$ means that relay node i is allocated to relay information for the source node k and $A_{k,i} = 0$, otherwise. Condition I in (28) means that at most one relay node can be allocated to a given source node and condition II means that each relay node can only relay information for one source node.

Since N and N' are limited, the optimal solution can be obtained through the following exclusive searching algorithm. Specifically, the destination node

- generates an $N \times N'$ matrix, $\mathbf{X}$, which consists of elements $x_{k,i} = \max\{I^{i}_{AF,k}, I_{D,k}\}$, $k = 1, \ldots, N$ and $i = 1, \ldots, N'$;
- searches for $x_{k^*,i^*} = \max_{\{k,i\}} x_{k,i}$ and allocates relay node i^* to source node k^*;
- deletes the k^*th row and the i^*th column from the matrix $\mathbf{X}$, and repeats the previous step until $\mathbf{X} = \mathbf{0}$, where $\mathbf{0}$ denotes an all zero matrix;
- broadcasts or unicasts the allocation results to all relay nodes and notifies the source nodes to initiate transmissions.

Obviously, exclusive searching is a centralized allocation algorithm, which needs to collect all $N \times N'$ channel state information and carry out $\min\{N, N'\}$ comparisons.

If channel gains are dominated by propagation path loss, clustering source and relay nodes into groups can reduce the computational complexity of the exclusive searching algorithm. The basic idea comes from the fact that the relay node, which locates far from the source node or the destination node, can hardly provide performance improvement by relaying. Therefore, it can be expected that the source nodes and their potential relay nodes may cluster together. One possible clustering algorithm is to divide the whole coverage area into slices, as shown in Fig. 3, where the number of slices equals 4. After clustering, for uniformly distributed source and relay nodes, only one quarter of the source and relay nodes are involved in each slice on average. Therefore, only $\frac{1}{4}N \times N'$ of the channel state information and matrices of $\frac{1}{16}N \times N'$ size are needed.

Even though clustering can reduce the computational complexity, the searching algorithm still requires to collect $\frac{1}{4}N \times N'$ channel states, which may be impractical in some cases and may induce large system overhead for information exchange. For practical implementation, in the following, an allocation algorithm, which contains some distributive property, is proposed. The algorithm consists of two parts: feasible set generation and relay node allocation.

- **Feasible set generation:**
  Before a source node initiates packet transmission to the destination, some hand-shaking signals have to be exchanged, such as request to send (RTS) and clear to send (CTS) in ad hoc networks. In the following, for explanation purpose, we borrow the names RTS and CTS to represent the hand-shaking signals before data transmissions. Upon hearing these hand-shaking signals, each relay node can estimate the channel gains from both the source and the destination nodes. According to the sufficient condition derived in the previous section, each relay node can make decision on its feasibility for a given source node by comparing its estimates of channel gains with the threshold. Such decision-making procedure is distributed and no extra information exchange is required. After that, all feasible relay nodes will report their node indices to the destination node, which can generate a feasible set for each source node.
- **Relay node allocation:**
  For networks with multiple source nodes, the source node with bad channel condition experiences low mutual information compared to those with good channel conditions. Therefore, to provide some fairness among different source nodes, higher priority should be given to the source nodes with bad channel conditions. This is equivalent to the criterion that maximizes the minimum capacity experienced by all the source nodes. Given the allocation priority among the source nodes, we can sort them and do the allocation one by one. Since there is no channel gain information at the destination for each source node, one of relay node is randomly selected from its feasible set.

Combining these two parts together, we can summarize the proposed user relaying algorithm as follows.

- When a source node has packets waiting for transmission, it sends an RTS to the destination node. This RTS can be heard by the destination node as well as all relay nodes in the source's transmission range. Each relay node can estimate the channel gain, β_{si}, from the source node.

- After receiving the RTS, the destination node estimates the channel gain, β_{sd}, from the source node and calculates the threshold C_s based on the estimation. The destination then feeds back an CTS as an acknowledgement, which includes the information on the threshold C_s. The relay nodes located in the transmission range of the destination node can receive this CTS and estimate the channel gain, β_{id}. Note that each relay node does not hold the information about the channel gains associated with other relay nodes.
- The relay nodes, who received both RTS and CTS, compare the channel gains with the threshold to determine its feasibility. The feasible relay nodes will notify the destination node of their existence.
- After receiving all notifications from the feasible relay nodes, the destination node generates a feasible set for each source node.
- The destination node sorts the source nodes in ascending order based on the estimates of β_{sd} and the relay node allocation is carried out for each source node.
- For a given source node, one of the relay nodes from its feasible set will be randomly chosen.
- The destination removes the selected source node's feasible set and removes the selected relay node from all feasible sets.
- The destination repeats the previous steps until there is no non-zero feasible set.
- The destination broadcasts the relay node allocation to all relay nodes and notify the source nodes to initiate transmissions.

In this algorithm, each relay node determines its participation individually and there is no need for the destination node to collect the information of channel gains. It can significantly reduce the system overhead resulting from the information exchange. However, since the final allocation decision is made by the destination node, this is the rationale for the term "semi-distributed".

V. Simulation Results

In this section, simulation results are presented to demonstrate the effectiveness of the proposed user relaying algorithms. The system performance evaluation focuses on the achievable average mutual information and computational complexity, measured as the average search time. We define the *search time* as the number of nodes searched during one execution of the search algorithm.

A. Simulation Parameters

Consider a single-destination relay network, which covers a circle area with a radius of 50 meters. The destination node is located at the center of the circle, while all other nodes are uniformly distributed in the covered area. Networks with both single source node and multiple source nodes are simulated.

The channel between any pair of nodes (i.e., source-relay, relay-destination, source-destination) introduces propagation path loss and fading. The propagation path loss is defined as the power law of the distance between the transmitting node and the receiving node, i.e.,

$$L_{ij} = d_{ij}^{-\gamma} \tag{29}$$

where L_{ij} and d_{ij} denote the propagation path loss and the distance between transmitting node i and receiving node j, respectively; and γ is the path loss exponent, which equals 2 in our simulation. The fading, χ_{ij}, is an i.i.d. Rayleigh random variable with a unit variance. Combining both the propagation path loss and fading, the channel gain between nodes i and j is given by

$$\beta_{ij} = |\alpha_{ij}|^2 = d_{ij}^{-\gamma}|\chi_{ij}|^2. \tag{30}$$

All channel gains vary slowly in time. The background noise is an i.i.d. complex Gaussian random variable. The transmission power is represented through SNR at the transmitter end, which is normalized by the variance of the background noise.

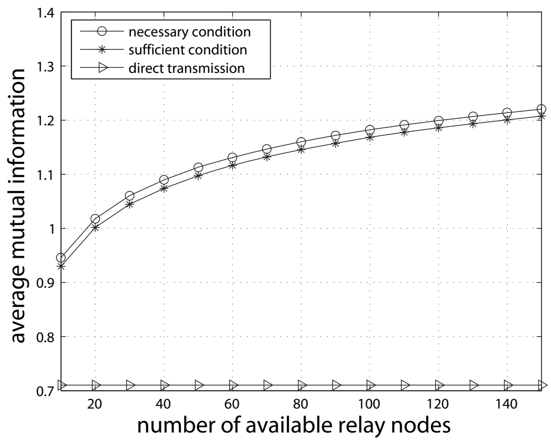

Fig. 4. Effects of the number of available relay nodes at $SNR = 30$dB.

B. Simulation Results and Discussions

We first evaluate the performance of a relay network in the single-source case. The source node is assumed to be located at the edge of the coverage area. It is the worst scenario since the source node has the maximum distance to the destination so that it has the largest propagation path loss and may have a maximum potential set of relay nodes based on either the necessary condition ($\mathbf{\Gamma}_n$) or the sufficient condition ($\mathbf{\Gamma}_s$). Optimal power allocation between the source node and the relay node is omitted first for simplicity, i.e., the total transmission power is assumed uniformly distributed.

Fig. 4 shows the effects of the number of available relay nodes on the system in terms of achievable average mutual information at $SNR = 30$dB. It can be seen that the achievable average mutual information increases as the number of available relay nodes or the relay node density increases. The performance improvement results from the fact that, with a higher probability, there exists one relay node at the best relay location when the relay node density increases. When the number of available relay nodes is sufficiently large, the performance improvement diminishes. The reason for this is that only one relay node can be selected for the given source node. Therefore, further increasing the relay node density can only marginally improve the presence of a best relay node.

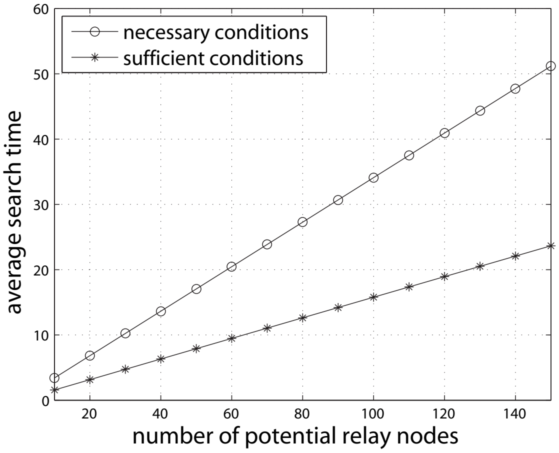

Fig. 5. Computational complexity of the user relaying algorithms at $SNR = 30$dB.

Fig. 4 also demonstrates that the user relaying algorithm based on the sufficient condition provides similar performance as that with the necessary condition. Throughput degradation is less than 0.016. It indicates that the feasible relay nodes omitted by the sufficient condition have negligible effects on the system performance.

Moreover, significant performance improvement is observed in Fig. 4 with user relaying over direct transmission. For instance, with 100 available relay nodes, the achievable average mutual information through user relaying is over 1.64 times more than that of direct transmission. Since the actual capacity equals the product of the maximum average mutual information and the system bandwidth, such performance improvement is significant for wideband communication networks.

Fig. 5 shows the computational complexity of two different user relaying algorithms in terms of the average search time required to find the best relay node. From the figure, a significant computational complexity reduction can be observed for the algorithm based on the sufficient condition. In our simulation, even for the maximum number of available relay nodes (i.e., 140), the average search time based on the sufficient condition is no more than 24 nodes, while the algorithm based on the necessary condition requires the search of more than 51 nodes. In addition, by comparing with the complete searching algorithm, which tests all available relay nodes, more than 83% of computational complexity can be saved by applying the sufficient condition. Since our simulation corresponds to the worst scenario, we can expect that the computational complexity can be further reduced for a situation where the source node is located inside the coverage area.

Fig. 6 shows the capacity performance improvement of the user relaying algorithms over the direct transmission with respect to different values of SNR. The total number of available relay nodes is set to 50. It can be observed that the performance improvement from the user relaying increases first and then decreases with the increase in SNR. It is because when SNR becomes very small, user relaying can only provide very limited cooperative diversity gain, as shown in (7). On the other hand, larger SNR results in higher thresholds of C_n and C_s, which reduce the size of the feasible relay node set. When SNR becomes sufficiently large, direct transmission becomes

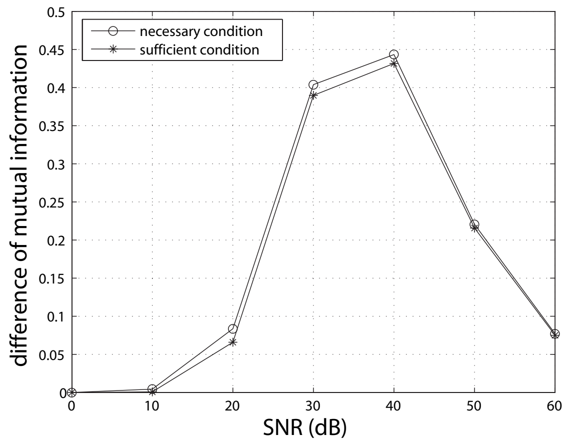

Fig. 6. Throughput difference between the user relaying algorithms and the direct transmission.

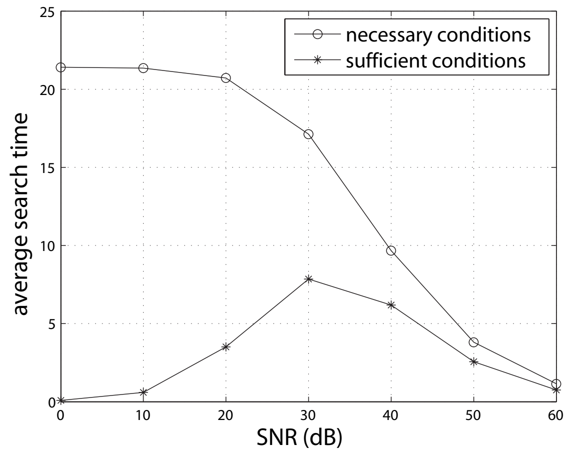

Fig. 7. Computational complexity of the user relaying algorithms with difference SNR.

the best choice to maximize system capacity. The simulation results also demonstrate that the user relaying algorithm based on the sufficient condition can provide similar performance as the one based on the necessary condition.

Fig. 7 shows the average search time of two user relaying algorithms as a function of SNR. For the algorithm based on the necessary condition, the average search time decrease as the SNR increases. The reason for this behaviour is because the threshold C_n is an increasing function of SNR. For the sufficient condition, the curve becomes concave with a maximum value less than 8 at $SNR = 30$dB. When SNR is very small, applying user relaying can only provide marginal capacity improvement over direct transmission, as shown in (7). Therefore, it is reasonable to shrink the size of $\mathbf{\Gamma}_s$ to only include the best relay nodes. On the other hand, when SNR becomes very large, direct transmission is the best so that the size of $\mathbf{\Gamma}_s$ tends to zero. Fig. 7 also demonstrates that the sufficient condition exhibits better average search time than that of the necessary condition. For instance, at $SNR = 30$dB, the average search time using the sufficient condition is nearly one third of those using the necessary condition.

Table I summarizes the effects of optimal power allocation on the performance improvement. In our simulation, the maximum performance improvement for the necessary condition and the sufficient condition is around 0.0858 and 0.0551,

TABLE I
AVERAGE MUTUAL INFORMATION W/ AND W/O POWER ALLOCATION

	NC w/ PA	NC w/o PA	SC w/ PA	SC w/o PA
SNR=0dB	0.0013	0.0013	0.0012	0.0012
10dB	0.0187	0.0162	0.0128	0.0128
20dB	0.2284	0.1922	0.1817	0.1745
30dB	1.2318	1.1460	1.1867	1.1316
40dB	3.1829	3.1258	3.1535	3.1153
50dB	5.8917	5.8709	5.8796	5.8661
60dB	9.0198	9.0130	9.0169	9.0119

NC: necessary condition; SC: sufficient condition; PA: power allocation.

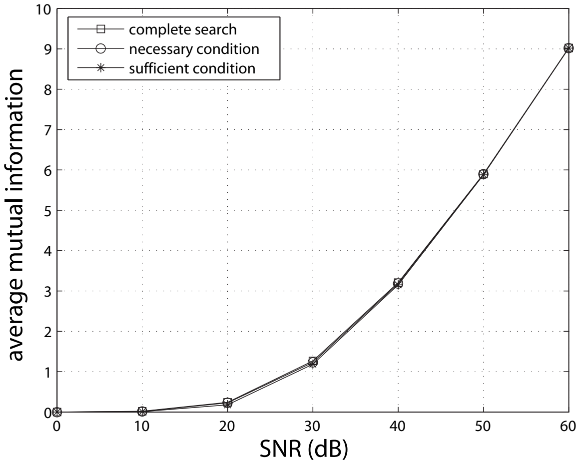

Fig. 8. Performance improvement from optimal power allocation.

respectively. Applying power allocation also introduces higher computational complexity. In practice, the tradeoff between the performance improvement and computational complexity should be balanced.

Fig. 8 shows the effectiveness of the decoupling introduced in Section III. The performance of the decoupling (i.e., power allocation + relay node selection based on the necessary condition or power allocation + relay node selection based on the sufficient condition) is compared with that of the complete searching algorithm. The figure shows that the three curves almost overlap with each other. It indicates that decoupling reduces the computational complexity, but only introduces negligible degradation on the overall capacity performance.

The performance of the relay network with multiple source nodes is shown in Fig. 9. The total number of nodes in the network is set to 101, which includes one destination node. One half of the 100 nodes are randomly chosen as the source nodes and the other half are considered as the relay nodes. It is observed from the figure that the performance is significantly improved by user relaying. The total capacity improvement over direct transmission is 13.4132 and 7.3283 for the exclusive searching algorithm and the semi-distributed algorithm, respectively. Since one relay node is randomly chosen from each source node's feasible set, the semi-distributed algorithm provides nearly one half of the performance gain achieved with the optimal exclusive searching algorithm.

VI. CONCLUSIONS

In this paper, user relaying algorithms, in terms of relay node selection and power allocation, are proposed for wireless relay networks. Threshold-based necessary and sufficient

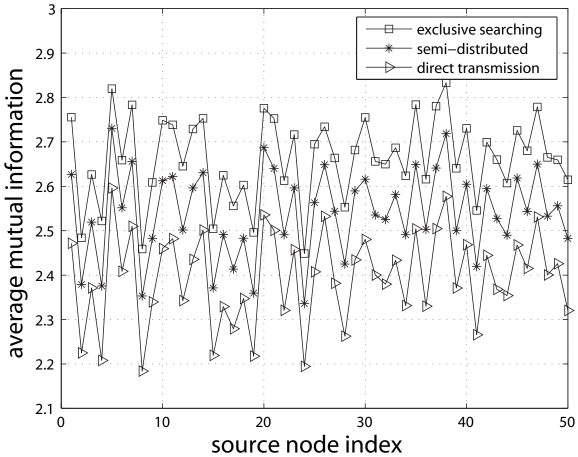

Fig. 9. Performance improvement from user relaying in the multiple-source network.

conditions have been derived for determining the feasibility of relay nodes. Based on the derived conditions, both centralized and semi-distributed user relaying algorithms have been proposed, which can provide nearly optimal capacity performance with significantly reduced computational complexity. The simulation results presented in Section V adequately demonstrate the effectiveness and efficiency of the proposed user relaying algorithms.

REFERENCES

[1] A. Raniwala, K. Gopalan, and T. Chiueh, "Centralized channel assignment and routing algorithms for multi-channel wireless mesh networks," *ACM SIGMOBILE Mobile Computing and Communications Review (MC2R)*, vol 8, No 2, pp. 50-65, Apr. 2004.

[2] R. Pabst, *et al.*, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Commun. Mag.*, vol. 42, no. 9, pp. 80-89, Sept. 2004.

[3] I. F. Akyildiz, X. Wang, and W. Wang, "Wireless mesh networks: a survey," *Computer Networks*, vol. 47, no. 4, pp. 445-487, Mar. 2005.

[4] A. Ghosh and D. R. Wolter, "Broadband wireless access with WiMax/802.16: current performance benchmarks and future potential," *IEEE Commun. Mag.*, vol. 43, no. 2, pp. 129-136, Feb. 2005.

[5] M. J. Lee, J. Zheng, Y. Ko, and D. M. Shrestha, "Emerging standards for wireless mesh technology," *IEEE Wireless Commun.*, vol. 13, no. 2, pp. 56-63, Apr. 2006.

[6] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3062-3080, Dec. 2004.

[7] N. C. Beaulieu and J. Hu, "A closed-form expression for the outage probability of decode-and-forward relaying in dissimilar Rayleigh fading channels," *IEEE Commun. Lett.*, vol. 10, no. 12, pp. 813-815, Dec. 2006.

[8] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity - Part I: system description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927-1938, Nov. 2003.

[9] J. Boyer, D. D. Falconer, and H. Yanikomeroglu, "Multihop diversity in wireless relaying channels," *IEEE Trans. Commun.*, vol. 52, no. 10, pp. 1820-1830, Oct. 2004.

[10] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 572-584, Sept. 1979.

[11] V. Mahinthan, J. W. Mark, and X. Shen, "A cooperative diversity scheme based on quadrature signaling," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, pp. 41-45, 2007.

[12] J. N. Laneman and G. W. Wornell, "Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inform. Theory*, vol. 59, pp. 2415-2525, Oct. 2003.

[13] G. Li and H. Liu, "Resource allocation for OFDMA relay networks," in *Proc Thirty-Eighth Asilomar Conference on Signals, Systems and Computers 2004*, vo. 1, pp. 1203-1207, Pacific Grove, CA, Nov. 7-10, 2004.

[14] I. Maric and R. D. Yates, "Bandwidth and power allocation for cooperative strategies in Gaussian relay networks," in *Proc. Thirty-Eighth Asilomar Conference on Signals, Systems and Computers 2004*, vol. 2, pp. 1907-1911, Pacific Grove, CA, Nov. 7-10, 2004.

[15] A. Bletsas, H. Shin, M. Z. Win, and A. Lippman, "Cooperative diversity with opportunistic relaying," in *Proc. IEEE WCNC'06*, pp. 1034-1039, Las Vegas, NV, Apr. 3-6, 2006.

[16] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE J. Select. Areas Commun.*, vol. 24, pp. 659-672, Mar. 2006.

[17] D. Gunduz and E. Erkip, "Outage minimization by opportunistic cooperation," in *Proc. IEEE Wirelesscom'05*, pp. 1436-1442, Maui, HI, June 13-16, 2005.

[18] V. Mahinthan, L. Cai, J. W. Mark, and X. Shen, "Maximizing cooperative diversity energy gain for wireless networks," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 2530-2539, July 2007.

[19] Y. S. Jung and J. H. Lee, "Partner assignment algorithm for cooperative diversity in mobile communication systems," in *Proc. IEEE VTC'06-Spring*, pp. 1610-1614, Melbourne, Australia, May 7-10, 2006.

[20] J. Cai, X. Shen, and J. W. Mark, "Robust channel estimation for OFDM wireless communication systems - an H_∞ approach," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 2060-2071, Nov. 2004.

Jun Cai (M'04) received the B.Sc. (1996) and the M.Sc. (1999) degrees from Xi'an Jiaotong University (China) and Ph.D. degree (2004) from University of Waterloo, Ontario (Canada), all in electrical engineering. From June 2004 to April 2006, he was with McMaster University as NSERC Postdoctoral Fellow. Since July 2006, he has been with the Department of Electrical and Computer Engineering, University of Manitoba, Canada, where he is an Assistant Professor. His current research interests include multimedia communication systems, mobility and resource management in 3G beyond wireless communication networks, and ad hoc and mesh networks. He is currently a holder of NSERC Associated Industrial Research Chair.

Xuemin (Sherman) Shen (M'97-SM'02) 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 the Associate Chair for Graduate Studies, Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research focuses on mobility and resource management in interconnected wireless/wired networks, UWB wireless communications systems, wireless security, and ad hoc and sensor networks. He is a co-author of three books, and has published more than 300 papers and book chapters in wireless communications and networks, control and filtering.

Dr. Shen serves as the Technical Program Committee Chair for IEEE Globecom'07, General Co-Chair for Chinacom'07 and QShine'06, the Founding Chair for IEEE Communications Society Technical Committee on P2P Communications and Networking. He also serves as a Founding Area Editor for *IEEE Transactions on Wireless Communications*; Editor-in-Chief for *Peer-to-Peer Networking and Application*; Associate Editor for *IEEE Transactions on Vehicular Technology*; *KICS/IEEE Journal of Communications and Networks*, *Computer Networks*; *ACM/Wireless Networks*; and *Wireless Communications and Mobile Computing* (Wiley), etc. He has also served as Guest Editor for *IEEE JSAC*, *IEEE Wireless Communications*, and *IEEE Communications Magazine*. Dr. Shen received the Excellent Graduate Supervision Award in 2006, and the Outstanding Performance Award in 2004 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 from the Faculty of Engineering, University of Waterloo. Dr. Shen is a registered Professional Engineer of Ontario, Canada.

Jon W. Mark (M'62-SM'80-F'88-LF'03) received the Ph.D. degree in electrical engineering from McMaster University, Canada in 1970. Upon graduation, he joined the Department of Electrical Engineering (now Electrical and Computer Engineering) at the University of Waterloo, became a full Professor in 1978, and served as Department Chairman from July 1984 to June 1990. In 1996, he established the Centre for Wireless Communications (CWC) at the University of Waterloo and has since been serving as the founding Director. He co-authored the recent text **Wireless Communications and Networking**, Prentice-Hall, 2003.

He is the recipient of 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, particularly in the areas of resource management, mobility management, and end-to-end information delivery with QoS provisioning.

Dr. Mark is a Life Fellow of the IEEE and has served as a member of editorial boards, including *IEEE Transactions on Communications*, *ACM/Baltzer Wireless Networks*, *Telecommunication Systems*, etc. He was a member of the Inter-Society Steering Committee of the *IEEE/ACM Transactions on Networking* from 1992-2003, and served as the SC Chair during 2002-2003.

Attahiru S. Alfa (M'00) is NSERC Industrial Research Chair of Telecommunications and Professor, Department of Electrical and Computer Engineering at the University of Manitoba, Winnipeg, Manitoba, Canada. Dr. Alfa carries out research in the areas of queueing and network theories with applications mostly to telecommunication systems. He has also applied these theories to manufacturing and transportation and traffic systems in the past. His current research interests are in the area of mobile communications, Internet traffic, stochastic models, performance analysis, network restoration, and teletraffic forecasting models. He has contributed significantly in the area of matrix-analytic methods for stochastic models. He has published in several journals, and most recently in Stochastic Models, Queueing Systems - Theory and Applications, Naval Research Logistics, *IEEE Journal on Selected Areas in Communications*, *IEEE Transactions on Vehicular Technology*, *Performance Evaluation*, *IEEE Transaction on Wireless Communications*, *Journal of Applied Probability*, *Advances in Applied Probability*, *Mathematics of Computations* and *Numerische Mathematik*. He received his B.Eng. in 1971 from the Ahmadu Bello University in Nigeria, M.Sc in 1974 from the University of Manitoba and Ph.D. in 1980 from the University of New South Wales, Australia. He belongs to the following organizations: APEGM, IEEE, and INFORMS.