Adaptive Exploitation of Cooperative Relay for High Performance Communications in MIMO Ad Hoc Networks

Shan Chu and Xin Wang
Department of Electrical and Computer Engineering
Stony Brook University, Stony Brook, New York 11794
Email: {schu, xwang}@ece.sunysb.edu

Abstract

With the popularity of wireless devices and the increase of computing and storage resources, there are increasing interests in supporting mobile computing techniques. Particularly, ad hoc networks can potentially connect different wireless devices to enable more powerful wireless applications and mobile computing capabilities. To meet the ever increasing communication need, it is important to improve the network throughput while guaranteeing transmission reliability. Multiple-input-multiple-output (MIMO) technology can provide significantly higher data rate in ad hoc networks where nodes are equipped with multi-antenna arrays. Although MIMO technique itself can support diversity transmission when channel condition degrades, the use of diversity transmission often compromises the multiplexing gain and is also not enough to deal with extremely weak channel. Instead, in this work, we exploit the use of cooperative relay transmission (which is often used in a single antenna environment to improve reliability) in a MIMO-based ad hoc network to cope with harsh channel condition. We design both centralized and distributed scheduling algorithms to support adaptive use of cooperative relay transmission. Our algorithm effectively exploits the cooperative multiplexing gain and cooperative diversity gain to achieve higher data rate and higher reliability under various channel conditions. Our scheduling scheme can efficiently invoke relay transmission without introducing significant signaling overhead as conventional relay schemes, and seamlessly integrate relay transmission with multiplexed MIMO transmission. We also design a MAC protocol to implement the distributed algorithm. Our performance results demonstrate that the use of cooperative relay in a MIMO framework could bring in a significant throughput improvement in all the scenarios studied, with the variation of node density, link failure ratio, packet arrival rate and retransmission threshold.

1. Introduction

The past a few years have witnessed a surge of interest in multiple-input multiple-output (MIMO) technology, which could potentially improve transmission reliability and provide higher raw data rates by utilizing multiple antennas at the transmitter and/or the receiver. Specifically, *multiplexing* takes advantage of the rich scattering environment to increase transmission capacity [1] and *diversity* effectively combats fading to enhance the transmission reliability [2], [3]. In order to exploit the benefits of MIMO technology, it is now being adopted in 802.11n [4] and also considered for ad hoc networks.

Some recent works have endeavored to apply MIMO techniques in ad hoc networks [5]–[13]. Although various MAC schemes have been designed to exploit the intrinsic features of MIMO to improve throughput and reliability, they may not be able to handle consecutive packet loss due to severe path loss, continuous deep fading or temporary topology changes and link breakages. Continuous packet retransmissions would lead to significant throughout reduction. Although beamforming can help improve transmission reliability, it compromises the potential multiplexing gain and hence reduces transmission rate. As an alternative to MIMO technique, recent efforts have been made to enable cooperative relay transmission to cope with channel degradation, with the assumption that network nodes have single antenna [14], [15]. One question to raise is: is it beneficial to adopt cooperative relay to facilitate transmission in a MIMO-based ad hoc network?

The introduction of cooperative relay transmission into a network where nodes are equipped with multiple antennas could bring in many potential benefits. It would not only allow joint exploitation of multiplexing gain of MIMO and cooperative diversity gain of relay transmission, but would also allow mitigation of many issues presenting in conventional relay transmissions. With the support of relay nodes, transmissions on MIMO links with harsh conditions or temporary breakages can possibly be bridged through relay links over source-relay-destination paths. Without being impacted by a poor link for a continuous time period, traffic can be scheduled more efficiently to avoid significant transmission delay and extra consumption of precious network resources. Besides, with careful relay selection, the channel quality of a relay link would be generally better thus allow for a higher rate, which reduces the cost of using relay transmission. Taking advantage of multi-packet transmission/reception capability enabled by MIMO technique, a relay node which has multiple antennas can overhear the transmission from a source while receiving its own packets, which avoids the need for the source to forward the packet explicitly to the relay node as in conventional cooperative transmission. Meanwhile, a relay node can simultaneously forward packet for others while transmitting its own packets.

Although there are significant benefits of using relay transmission in a MIMO ad hoc network, there are also big challenges in efficiently selecting and triggering cooperative relay transmissions, especially in concert with multi-user-based MIMO transmissions in a meshed network environment. Without a properly designed strategy, the use of relay would cost much more transmission time and bandwidth instead of supplementing the spatial multiplexing transmission.

In this paper, our focus is to design algorithms and MAC schemes that adaptively use cooperative relay in MIMO-based ad hoc networks to further improve transmission reliability and throughput. The main contributions of this paper are as follows.

- We formally formulate the problem and provide a centralized algorithm;
- We practically divide the problem into two phases and provide simple but effective distributed scheduling algorithms that seamlessly incorporate the use of cooperative relay into MIMO transmission;
- We propose a simple relay scheme to formulate relay set and invoke relay transmission without extra signaling overhead;
- We design an efficient MAC protocol to support our distributed algorithm.

The rest of the paper is organized as follows. We discuss the related work in Section 2 and introduce the motivation of our work in Section 3. We formulate the problem and propose a centralized algorithm in Section 4. We then present our scheduling algorithms to support seamless use of cooperative relay with multi-user-based MIMO transmission in an ad hoc network in Section 5, and provide more details about relay operation and MAC protocol design in Section 6. Finally, we study the performance through simulations in Section 7 and conclude the paper in Section 8.

2. Related Work

In recent years, many efforts have been made to support MIMO transmission in ad hoc networks. In [5], spatial diversity is explored to combat fading and achieve robustness. A centralized algorithm is presented in [6] to solve the joint routing, scheduling and stream control problem subject to fairness constraints in mesh networks with MIMO links. Layered space-time multiuser detection and its role in PHY-MAC cross-layer design are analyzed in [7]. In [8], spatial multiplexing with antenna subset selection for data packet transmission is proposed. The optimization considerations for MAC layer design in ad hoc networks with MIMO links is discussed in [9], and [11] exploits the benefits of using multiple antennas to achieve flow-level QoS in multi-hop wireless networks. In [13], an opportunistic and cooperative multiplexing scheme is proposed to better exploit spatial/multiuser diversity to improve transmission capacity and support different traffic demands in the network. However, none of these solutions considers the potential benefits of using cooperative relay in MIMO-based ad hoc networks. Though cooperative diversity has been extensively studied theoretically [14], there are limited work that investigate its practical network implementations. In [15], the authors proposed relaying strategies to increase the system reliability and the work in [16] tries to emulate the function and achieve the transmit diversity gain of using space-time codes in a distributed manner through node cooperation without the use of multi-antenna arrays. A multi-layer approach for exploiting virtual MISO links in ad hoc networks is presented in [10] and an optimal relay assignment is discussed in [17]. In [18], the authors analytically considers a general multiple-antenna network with multiple relays in terms of the diversity-multiplexing tradeoff. In [12], retransmission diversity through node cooperation is investigated in specific homogeneous omni-directional and smart antenna networks. Cooperative spatial multiplexing is systematically implemented with hybrid ARQ in [19], however, it lacks a detailed algorithm and protocol to specifically enable cooperative transmission which is generally very challenging to achieve in a dynamic network. Our work distinguishes itself from the aforementioned work in that it adaptively adopts relay forwarding with cooperative MIMO multiplexing to significantly improve the throughput while supporting transmission reliability.

3. Background and Motivation

With nodes equipped with multiple antennas, multiple data streams may be transmitted between a node pair through *spatial multiplexing*, while *spatial diversity* may be exploited to improve transmission reliability. Specifically, when channel information is available, a subset of antennas that can transmit signals at better quality can be selected for transmissions through *selection diversity*, which is shown to outperform space-time coding [20]. Instead of only allowing multiplexed transmission between a pair of nodes as in traditional MIMO scheme, in this work, we consider *cooperative MIMO multiplexed transmission* in which multiple nodes can simultaneously transmit to a receiver that has multiple antennas, and a sender with multiple antennas can also transmit multiple streams to a set of nodes. Therefore, a group of nodes can form *virtual MIMO array* [21].

This framework allows the exploration of multi-user diversity and antenna selection diversity to further improve capacity and reliability. The transmitter nodes and the antenna to use from a node are opportunistically selected based on the channel conditions between different nodes and antennas. These diversity techniques, however, are insufficient when the channel condition is extremely weak, the existence of correlated fading between a sender and receiver pair, or the distance between a node pair changes as a result of temporary topology change. If the channel degradation is short-term, it would be inefficient to change the transmission path immediately. Although schemes such as beamforming could be used between the transmission pair which has severe channel condition, it may prevent concurrent transmissions from the same or other nodes and compromise the potential throughput gain of the network that could be achieved with multiplexed transmissions. Also, when the distance between two nodes is too long or the channel is too weak, even beamforming is not enough.

In order to alleviate the problem of data rate reduction and excessive queuing delay caused by severe channel condition and/or link breakage as a result of temporary network topology change, in this work, we propose to *adaptively use cooperative MIMO transmission and cooperative relaying* when direct transmission cannot be successfully pursued. There are some unique benefits by taking advantage of both techniques: 1) Different from literature work which exploits cooperative diversity in a single antenna case only to improve transmission quality, in the proposed work, the relay transmissions coordinate with the transmissions in a neighborhood and take advantage of *cooperative multiplexing* to improve the overall network throughput; 2) Instead of simply delaying the transmissions of the relay nodes, which is often the case in the conventional cooperative diversity study, a relay node can be given more opportunities for transmissions by transmitting multiple streams simultaneously with use of multiple antennas, or having a higher transmission probability driven by our priority based scheduling as the priority of a relay node increases when its packets experience more delay due to relay transmissions; 3) The direct transmissions and relayed transmissions are performed independently, and a receiver node takes advantage of multiple antennas to decode transmissions from multiple streams without requiring synchronization at the symbol level between neighboring nodes as in conventional cooperative diversity schemes; 4) With multi-packet reception capability, a relay node can obtain the packet to be relayed through overhearing during its own data receiving when the sender attempts for direct transmission initially. In Fig. 1(a), R receives the relay packet as an interference stream while it is receiving data stream from Q. It can also simultaneously transmit to Q when it serves as a relay node to transmit the relay packet to D, as shown in Fig. 1(b).

With use of coded cooperation, the network performance can be further improved. As our focus is to investigate the benefit and strategy of incorporating relay into multiplexed MIMO transmission, we consider decode and forward cooperative strategy here for simplicity.

4. Problem Formulation

In this section, we first describe the system model. Then we formally formulate the problem and provide a centralized algorithm.

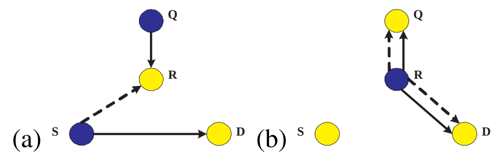

Figure 1. An illustration of cooperative relay transmission.

To enable concurrent many-to-many stream transmission, our MAC design is TDMA based, in which the time domain is divided into transmission durations (TD). A TD consists of several time slots with fixed length and covers one round of control signal exchange and data frame transmission. The data transmission rate within a frame can vary based on the channel condition. For a channel with higher quality, more efficient coding can be used to encode the symbols at a higher rate. The total transmit power at each node is considered to be fixed, while the transmit power of each antenna is different when a node uses a different number of antennas for transmission.

There is a buffer queue at each node where data packets are stored. For a packet p_i, a parameter called *priority* $\mathscr{P}(p_i)$ is used to capture both its service type and queuing delay. For the convenience of calculation, $\mathscr{P}(p_i)$ is measured in the unit of TD. A possible way to integrate both factors into the priority calculation is to equate the service priority of p_i to an initial value of $\mathscr{P}(p_i)$ in terms of TD, and $\mathscr{P}(p_i)$ increases as the queuing time of p_i increases. A higher value of $\mathscr{P}(p_i)$ indicates that the packet p_i has a higher priority.

The transmissions of packets are organized as *streams*. For spatial multiplexed transmission, a stream s is defined to be an independent data flow transmitted from an antenna of a transmitter node to a receiver node and identified by a triplet $s = (I_t, I_r, I_{ant})$, where $I_t/I_r/I_{ant}$ is the index of the transmitter/receiver/antenna that involves in the transmission of the stream. Given the signal to noise and interference ratio (SINR) $\rho_{I_r}(s)$ for stream s at the receiver node, the data rate of s can be calculated as $\mathscr{R}(s) = \log(1 + \rho_{I_r}(s))$. In a practical system, a receiver can feedback its estimated $\rho_{I_r}(s)$ to a transmitter which then decides the actual data rate it uses for transmission through looking up a pre-set table for instance.

As the complete information about future traffic is unavailable, it is a practical option to schedule the transmission of packets in each transmission duration (TD) considering the existing traffic and queueing delay. Suppose there is a set of nodes $N = \{n_1, n_2, \ldots, n_{N_n}\}$ in the network, and there are N_p packets waiting for transmission which are contained in the set $P_{pkt} = \{p_1, p_2, \ldots, p_{N_p}\}$. A node n_j has an antenna array of size N_j^{ant}. We consider half-duplex transmission, with which a node cannot be the transmitter and receiver at the same time. In a TD, a subset of nodes, denoted as T, are selected as transmitter nodes.

Denote the set of neighboring nodes of node n_j as $\mathscr{V}_j$. After a direct transmission of a packet p_i from s_i to d_i, nodes in $R_i = \{n_r | s_i \in \mathscr{V}_r, d_i \in \mathscr{V}_r, n_r \in N \setminus T\}$ overhear the transmission and store the packet in their own buffers. These nodes become candidate relay nodes for p_i. The packet p_i becomes available to nodes in $R_i \bigcup \{s_i\}$, which store the packet with the consistent priority. R_i is updated to include more qualified relay nodes whenever there is any direct transmission of p_i. Consider $i = 1, \ldots, N_p$, and $j = 1, \ldots, N_n$. Let the parameter $x_j \in \{0, 1\}$ denotes the transmitter node assignment in the current TD. If node n_j is selected as a transmitter node, $x_j = 1$, otherwise $x_j = 0$. $y_{ij} \in \{0, 1\}$ is used to denote the association of a packet and a transmitter node in the current TD. If a packet p_i is transmitted from node n_j, $y_{ij} = 1$, otherwise, $y_{ij} = 0$. The assignment of a packet to a specific antenna of a transmitter is represented by $a_{ijk} \in \{0, 1\}, (k = 1, \ldots, N_j^{ant})$, where $a_{ijk} = 1$ if and only if packet p_i is assigned to be transmitted from the k-th antenna of node n_j. Note that if $a_{ijk} = 1$, the transmission rate of packet p_i depends on the channel condition of the stream $s(i) = (j, d_i, k)$ and the interference at node d_i when receiving the stream, denoted as $\mathscr{I}(d_i)$. Therefore, the rate of stream $s(i)$ is denoted as $\mathscr{R}(s(i), \mathscr{I}(d_i))$.

Following the scheduling framework in [22], our scheduling aims to jointly optimize both data rate and priority. To exploit the benefits enabled by both multiple antennas and relay nodes, we formulate the scheduling problem as follows to capture the features of MIMO transmissions and conditions of relay transmissions:

$$\max \sum_{i=1}^{N_p} \sum_{j \in R_i \bigcup \{s_i\}} \sum_{k=1}^{N_j^{ant}} x_j (1 - x_{d_i}) y_{ij} a_{ijk} \mathscr{R}(s(i), \mathscr{I}(d_i)) \mathscr{P}(i); \tag{1}$$

$$\sum_{j \in R_i \bigcup \{s_i\}} x_j y_{ij} \le 1, i = 1, 2, \ldots, N_p; \tag{2}$$

$$x_j \sum_{i=1}^{N_p} y_{ij} a_{ijk} \le 1, j = 1, \ldots, N_n, k = 1, \ldots, N_j^{ant}; \tag{3}$$

$$(1 - x_j) \sum_{i=1}^{N_p} \sum_{m \in \mathscr{V}_j} \sum_{k=1}^{N_m^{ant}} x_m y_{im} a_{imk} \le N_j^{ant}, j = 1, 2, \ldots, N_n; \tag{4}$$

$$x_j, y_{ij}, a_{ijk} \in \{0, 1\}.$$

Constraint (2) ensures that a packet p_i is assigned to at most one transmitter node among all the candidate ones (including the source node s_i and candidate relay nodes in R_i) to avoid redundant transmission. Constraint (3) represents the *transmitting constraint* that an antenna k at a transmitter n_j can only accommodate the transmission of at most 1 stream in a TD. Constraint (4) provides the *receiving constraint* to model the impact of interference at the receiver end of a MIMO link, where the total number of receiving streams (data streams plus interference streams) at a receiver node n_j is restricted to be no more than its number of antennas in order to decode the receiving data packet. With this formulation, the nodes without packets will have the priority set to 0 and not be scheduled, while the ones with worse channel condition but higher priority will be transmitted.

So far, we formulate the problem of cooperative transmission with relays in a MIMO ad hoc network as an integer programming problem with objective function in (1) subject to constraints (2)(3)(4). In Algorithm 1, we propose a centralized scheme to schedule packet transmissions in a TD. As the interference streams which can transmit simultaneously with stream i are unknown before scheduling is made, it makes the accurate determination of $\mathscr{R}(s(i), \mathscr{I}(d_i))$ difficult. On the other hand, as the transmission rate is only used as a guidance to select the streams that potentially support higher throughput for transmissions, there is no need to know the accurate transmission rate at scheduling time. Therefore, for each candidate stream, we estimate the transmission rate based on the received signal strength of stream $s(i)$ calculated based on the knowledge of the channel condition of the stream, and conservative estimation of the interference at its receiver d_i which is calculated based on the received signal strength of $N_{d_i}^{ant} - 1$ strongest streams around d_i. In a practical transmission, with channel condition from all the potential transmitters estimated in advance, the majority of interference can be canceled, and the actual transmission rate can be higher. To facilitate scheduling, a weight $w(ijk)$ is introduced to represent the benefit achieved with the transmission of packet p_i from transmitter n_j using antenna k, and the set W consists of the weights of all candidate streams. The algorithm greedily schedules a packet p_i^* to transmit from antenna k^* of transmitter node j^*, which has the highest weight among all the candidate ones and guarantees the constraints (3)-(4). P is the set of scheduled packets and T contains all selected transmitters. On line 12, all the candidate streams that have transmission conflict with the scheduled stream $s = (j^*, d_{i^*}, k^*)$ are removed from the set W, including the ones that have the node n_{j^*} as the receiver, have $n_{d_{i^*}}$ as the transmitter, or have node n_{j^*} as the transmitter but are associated with the antenna k^*. A packet may be queued at multiple candidate transmitting nodes (i.e., source and candidate relay nodes). To avoid repetitive transmission of a packet and satisfy constraint (2), all other candidate streams for the selected packet p_{i^*} are also removed from W after p_{i^*} is successfully scheduled in the current TD. The algorithm then checks if packets are correctly received at destinations on lines 18-19, and successfully received packets are removed from the packet set P_{pkt}. For any incorrectly received packet p_i, its candidate relay list R_i is updated to add in nodes that are within the range of both the source and destination of p_i and have correctly overheard the direct transmission, as on lines 21-23, so that nodes in R_i would assist in the transmission of p_i in the following TDs.

5. Packet Scheduling with Relay Transmission

In order to achieve optimum system performance, it is essential for a scheduling scheme to determine the set of nodes that serve as the transmitters and the packets to be transmitted in a transmission duration, and assign them to the appropriate antennas for transmissions. Our scheduling

Algorithm 1 Centralized Scheduling

Initialize:
$W \Leftarrow \oslash$, $T \Leftarrow \oslash$, $P \Leftarrow \oslash$, $y_i \Leftarrow 0$, $x_j \Leftarrow 0$, $a_{ijk} \Leftarrow 0$, $\forall i, j, k$, update P_{pkt} to include new packets
for $\forall p_i \in P_{pkt}$ **do**
  for $\forall n_j \in R_i \bigcup \{s_i\}$ **do**
    $w(ijk) \Leftarrow \mathscr{R}(s(i), \mathscr{I}(d_i))\mathscr{P}(i), \forall k \in \{1, \ldots, N_j^{ant}\}$
    $W \Leftarrow W \bigcup \{w(ijk)\}$
  end for
end for
Scheduling:
while $W \neq \oslash$ **do**
  $(i^*, j^*, k^*) = \arg\max_{\{i,j,k\}} W$, the corresponding destination node is d_{i^*}
  if Selecting stream (j^*, d_{i^*}, k^*) satisfies (3) for n_{j^*} and (4) for all nodes in $\mathscr{V}_{j^*}$ **then**
    Schedule the stream (j^*, d_{i^*}, k^*), $y_{i^*} \Leftarrow 1$, $x_{j^*} \Leftarrow 1$, $a_{i^*j^*k^*} \Leftarrow 1$, $P \Leftarrow P \bigcup \{p_{i^*}\}$, $T \Leftarrow T \bigcup \{n_{j^*}\}$
    $W \Leftarrow W \setminus \{w(ijk) | \forall d_i = j^*\} \bigcup \{w(ijk) | \forall j = d_{i^*}\} \bigcup \{w(ijk) | j = j^*, k = k^*\} \bigcup \{w(ijk) | i = i^*\}\}$
  else
    $W \Leftarrow W \setminus w(i^*j^*k^*)$
  end if
end while
Relay Set Update:
for $\forall p_i \in P$ **do**
  if p_i is correctly decoded at n_{d_i} **then**
    $P_{pkt} \Leftarrow P_{pkt} \setminus \{p_i\}$
  else
    for $\forall n_m \in \{n_r | n_r \in \mathscr{V}_{s_i} \bigcap \mathscr{V}_{d_i}, \sum_k a_{is_ik} \geq 1, n_r \in N \setminus T\}$ **do**
      if p_i is correctly decoded at n_m **then**
        $R_i \Leftarrow R_i \bigcup \{n_m\}$
      end if
    end for
  end if
end for

algorithm fully exploits the multiplexing gain enabled by cooperative MIMO transmission and diversity gain enabled by cooperative relay transmission for overall higher system performance.

From the problem formulation in Section 4, it is clear that the scheduling problem has to determine the values of the three parameter set: $\{x_j\}$, $\{y_{ij}\}$, and $\{a_{ijk}\}$ to assign a packet to an appropriate transmitter antenna in order to maximize the total weighted rate of the network. In a practical half-duplex network, it is reasonable to divide the problem into two parts: transmitter selection and stream allocation. In the first phase, a set of nodes are selected as transmitter nodes, and for each selected node, it needs to determine the number of packets to transmit in the current transmission duration. Thus the values of $\{x_j\}$ and $\{y_{ij}\}$ are determined. The decision in our scheduling is made based on the transmission priority of the packets in queue, and the antenna constraints of the transmitter nodes and receiver nodes. In the second phase, each selected transmitter node needs to assign its packets to appropriate antennas for transmission based on the number of streams it is allowed to transmit, the priority of the packets, and the channel conditions. Thus, the value of $\{a_{ijk}\}$ is determined. In the next two subsections, we introduce the problem and algorithm for each scheduling phase.

5.1. Determination of Transmitter Nodes and the Number of Transmission Streams

Instead of randomly selecting the transmitter nodes in a TD, in this phase, we propose a *priority-based self-selection* strategy with which an active node self-determines if it can serve as the transmitter and the number of streams to transmit based on the priority of its packets, its transmitter constraint and the decoding constraints of its neighbors. A candidate relay node incorporates the relay packet with its own transmission and participates in the transmitter selection process.

As the selection is performed at the beginning of each TD before any transmissions, the rate information for candidate streams is unavailable. The transmitter node assignment and the number of streams are thus determined with the goal of optimizing the overall priority performance, and the goal of rate optimization is addressed later in the stream allocation phase.

5.1.1. Distributed Transmitter Node Selection. Let Q_j denote the packet queue at node n_j, where original packets and relay packets are sorted in a descending order of their priorities. Let N_j^0 be the proposed number of transmission streams, obviously $N_j^0 = \min\{N_j^{ant}, |Q_j|\}$. Parameter U_j is defined to be the priority of the head-of-the-line packets in node n_j's queue, i.e., $U_j = \sum_{l=1}^{N_j^0} \mathscr{P}_{p_{j,l}}$, which is used as the priority of n_j for scheduling.

In order to avoid unnecessary channel measurement and message processing at a receiver, we introduce a probability P_j^{TX}, below which an active node n_j can be selected as a transmitter node. Suppose n_m, a neighboring node of n_j, has N_m^{active} neighboring nodes and can decode N_m^{dec} concurrent streams, which can be obtained from periodic Hello messages sent in the two-hop neighborhood of n_j at the network layer. If the average number of streams from a single transmitter node around a receiver n_m is known and denoted as $\bar{N}_{\mathscr{V}_m}^{allo}$, in order to not exceed its decoding capacity, n_m generally only allows $\widetilde{N}_m = N_m^{dec}/\bar{N}_{\mathscr{V}_m}^{allo}$ nodes among its N_m^{active} neighbors to transmit in a TD. That is, each of the nodes around n_m is allowed to have a probability of $N_m^{dec}/(\bar{N}_{\mathscr{V}_m}^{allo} N_m^{active})$ to serve as the transmitter. As $\bar{N}_{\mathscr{V}_m}^{allo}$ is hard to know before scheduling is performed, a node can at most have a probability of N_m^{dec}/N_m^{active} to serve as the transmitter. The parameter P_j^{TX} of n_j can then be calculated as follows to consider the decoding capability of all its neighboring receiver nodes:

$$P_j^{TX} = \min_{m \in \mathscr{V}_j} \Big(N_m^{dec}/N_m^{active} \Big). \tag{5}$$

Instead of only considering the decoding capability of the selected receiver nodes which is not available at the selection time, our selection considers the decoding capability of all the neighboring nodes and is more conservative.

With this calculation, when there is only a small number of nodes around each receiver, there is a possibility that all the nodes within a neighborhood are selected as the transmitters. For example, if the network has only two nodes and each node can decode up to four streams, both nodes may be selected as transmitters and it is not possible to complete the transmission. To avoid this problem, when $P_j^{TX} \geq 1$, the value of P_j^{TX} is replaced with $P_j^{TX} = \max_{m \in \mathscr{V}_j} (N_m^{active}/(N_m^{active}+1))$, so that at least one node will be kept as the receiver.

The priority of a node can be attached with periodic Hello messages sent at the network layer, and updated with the data packets sent. The priority of the active nodes not having packets sent in a TD can be predicted as time moves forward. A node n_j can then record the maximum priority U_j^{max} and the minimum priority U_j^{min} of all the N_j^{active} active nodes in its neighborhood and itself, and also calculate the average priority $\bar{U}_j$ as $\bar{U}_j = (\sum_{m=1}^{N_j^{active}} U_m + U_j)/(N_j^{active}+1)$.

To avoid extra signaling and control overhead, an active node n_j *self-decides* if it should be selected as a transmitter node by calculating an index number r_j^{TX} as follows:

$$r_j^{TX} = \begin{cases} (\bar{U}_j - U_j)/(U_j^{max} - U_j^{min}) + \gamma_j & \text{if } U_j^{max} \neq U_j^{min} \\ \gamma_j & \text{if } U_j^{max} = U_j^{min} \end{cases} \tag{6}$$

where the parameter γ_j is uniformly distributed in the range $[0, 1]$ and randomly generated by a node j in each transmission duration (TD) to provide some fairness among nodes. The factor $(\bar{U}_j - U_j)/(U_j^{max} - U_j^{min})$ is used to give the higher priority node a larger probability for transmission. In a TD, if $r_j^{TX} < P_j^{TX}$, node n_j is selected as a transmitter node; otherwise, it has no right of transmission. Our transmitter selection algorithm gives preference to a node with a higher service priority and/or a larger load and hence longer delay, and thus supports load balancing while ensuring certain fairness.

Note that in this phase relay packets and original packets are treated equally, and the value of $\{x_j\}$ is determined.

5.1.2. Distributed Determination of the Number of Streams. Through the procedure described next in Section 6, a receiver node estimates the total number of candidate streams it may receive N_j^{rec} and broadcasts it together with the number of streams it is able to decode N_j^{dec}. These two parameters are used at a transmitter node to determine the actual number of transmission streams it is allowed to transmit.

Denote the set of receiver nodes within the transmission range of a transmitter node n_j as X_j^{rc}. In order to ensure all the receiver nodes in its neighborhood to have high probability of meeting degree constraint, n_j constrains its number of sending streams to a number N_j^{allo} as follows:

$$N_j^{allo} = N_j^0 \min_{m \in X_j^{rc}} \left(\frac{N_m^{dec}}{N_m^{rec}} \right). \tag{7}$$

Note that the value N_j^{allo} may be a fractional number. To achieve a higher accuracy in calculating N_j^{allo} than using simple rounding, let $N_{j,0}^{allo} = N_j^{allo} - \lfloor N_j^{allo} \rfloor$. If $N_{j,0}^{allo} > 0$, generate a random variable β_j uniformly distributed in $[0, 1]$. If $\beta_j \leq N_{j,0}^{allo}$, $N_j^{allo} = \lfloor N_j^{allo} \rfloor + 1$; otherwise, $N_j^{allo} = \lfloor N_j^{allo} \rfloor$.

For a packet p_i, there may be several candidate relay nodes. It would waste network resource if several nodes forward the same packet. Our scheduling scheme naturally selects the forwarding nodes based on the relevant priority of the to-be relayed packet and the priorities of the other packets of a relay node. After this self selection process, there are still the possibility that some relay nodes choose the same TD to forward p_i. To further reduce the chance of unnecessary relay forwarding, when the destination receives multiple relay transmission requests, it selects the relay node with the best channel condition to forward the packet. The rest of the requesting relay nodes can use the slot to send other packets.

5.2. Allocation to antennas

In this phase, the first N_j^{allo} data packets in the queue of node n_j are allocated to N_j^{allo} out of N_j^{ant} antennas for transmission. The packets may have different destination nodes thus varied link loss, and the spatial channels from different elements of the antenna array undergo independent fading. As discussed in [13], the data rate can be improved by opportunistically allocating the packets to transmitted antennas. Moreover, with channel information available at transmitters' side, selection diversity is shown to outperform space-time coding in improving the link reliability [20]. With the goal of maximizing transmission rate, the stream allocation problem is essentially a bipartite maximum matching problem.

Construct a graph $G = (V_1 \cup V_2, E)$ for a transmitter node n_j. V_1 denotes the set of packets to be allocated to antennas and V_2 denotes the set of transmitting antennas of n_j. Thus $|V_1| = N_j^{allo}$ and $|V_2| = N_j^{ant}$. Form an edge (v, u) between v and u where $v \in V_1$ and $u \in V_2$, and the weight of the edge is $w_{vu} = \mathscr{R}(v, u)$. Here $\mathscr{R}(v, u)$ is the rate of the stream to transmit a packet represented by node v to its destination node through the antenna represented by node u, which is estimated through signal exchange as discussed in Section 6. If $|V_1| \neq |V_2|$, add dummy nodes to make $|V_1| = |V_2|$ and the edges connected to a dummy node has weight 0.

By solving the maximum weight matching problem formulated above (i.e. using successive shortest path algorithm [23]) and then deleting the dummy nodes and edges connected to them, the optimum solution of the allocation is derived.

6. Protocol Design

In the previous section, the scheduling is performed in each transmission duration to determine the transmission schedule of the packets, including original packets and relay packets, in the queue of each node. However, the details about cooperative relay transmission, i.e. how to maintain the queue to store relay packets, how to trigger and enable a relay node to transmit relay packets have not been addressed yet. In this section, we propose the protocol to facilitate cooperative relay transmission in a MIMO-based ad hoc network and implement the distributed scheduling algorithm described in section 5.

6.1. Relay Operations

There are several challenges arising in integrating the cooperative relay transmission with the cooperative MIMO multiplexing transmission scheme.

6.1.1. Finding candidate relay nodes. In a conventional relay strategy, a source often broadcasts a relay request explicitly, and waits for replies from the potential relay nodes. This process not only introduces extra signaling overhead, but also adds in delay for relay transmission. Instead, the process of finding candidate relays in our scheme is automatically performed at qualified nodes without involving the source and destination of a packet. Specifically, a node r_i determines its role of being a candidate relay node of a packet p_i which is targeted to d_i when successfully receiving the packet from its sender s_i, either because r_i is idle or because r_i could decode p_i when receiving its own packet with its multi-packet reception capability. If the destination of the data packet p_i is also in r_i's neighbor list, r_i temporarily stores p_i in its buffer with the current priority of p_i. If p_i is successfully received by d_i, r_i removes p_i from its buffer; otherwise, the priority of p_i is updated as its buffering time in r_i increases. In a dense network, to avoid excessive buffering, a node may only buffer a packet with certain probability, or a sender could tag the packets that may need relay.

6.1.2. Triggering of relay transmission. Instead of explicitly invoking relay transmission, in our scheme, triggering of relay transmission and selection of relay node is incorporated with normal packet scheduling. If a failed direct transmission is detected, i.e. a candidate relay r_i receives packet p_i from s_i but does not receive the successful reception acknowledgement from d_i in the same TD, r_i immediately moves the relay packet p_i from buffer to its MAC queue, and treats it as a normal packet waiting for transmission. The node then serves as a relay node in the following TDs. Note that the source node and relay nodes of p_i may intend to transmit it in the same TD, if p_i happens to be a head-of-the-line packet in all of their queues. In order to reduce the chance of concurrent transmission, the targeted receiver node counts the number of transmission requests for the same packet. The receiver selects the node with the best channel condition to serve as the packet sender and broadcasts the selection.

6.1.3. Constraining the delay of relay transmission. To avoid excessive traffic increase and occupation of network resource, a retransmission threshold F is introduced that a packet is dropped if F TDs has elapsed since its first direct transmission. To ensure that the source node and all candidate relay nodes have a consensus on the packet transmission status, a packet transmitted from its source node is attached with a time-stamp indicating the current elapsed time since its initial transmission, so that candidate relays can record this stamp and update it as the queuing time increases.

6.1.4. Broadcast of packet reception status. The information about successful or failed reception of a packet is usually broadcast through ACKs. However, as all receivers in a TD send ACK simultaneously as described in Sections 6.2, only nodes that are not receivers in the current TD can receive the ACKs. As a candidate relay node may either serve as a transmitter or a receiver in a TD, it is necessary to inform all of them about the reception status, so that successfully received packets can be removed while unsuccessfully received packets can have their priority updated. In addition, if the channel condition between the destination and source is very poor, a source may not be able to get the ACK. To address both issues, an extra ACK phase is introduced into the protocol, during which the information included in the first ACK are rebroadcast by non-receiver nodes in the current TD. To differentiate between the two ACK messages, they are named ACK-I and ACK-II respectively. In the proposed MAC scheme, the data transmission can be in burst, so the overhead of ACK signaling is relatively small.

6.1.5. Rate determination. As described in the protocol, both transmitter nodes and receiver nodes are able to estimate the full channel condition matrix through training sequences. Also, a receiver node can estimate the interference and noise around it, and announce this information to the corresponding senders. With the channel matrix and the interference and noise at the receiver, a transmitter can determine the rate to use for transmission. If a packet is scheduled for its first direct transmission and the link to its destination is estimated to be severe, the source node uses a default moderate transmission rate for its transmission, so as to increase the chance of having some relay node successfully receive the packet as well as avoid wasting the transmission opportunity in the current TD. Note that a packet transmission is canceled if a sender node could not receive response from the receiving after sending an initial handshaking signal. A source node may even give up its transmission towards a particular receiver if the transmission fails continuously over a period of time, i.e. longer than $3 \times F$ TDs, and look for an alternative path to the destination.

6.2. Protocol Details

Based on the above operations, we propose a TDMA-based MAC protocol to support the cooperative relay transmission in a MIMO-based ad hoc network. A time frame is divided into five phases with different transmission duration, namely RTS, CTS, DATA, ACK-I, and ACK-II. Note that slot synchronization is currently achievable in the IEEE 802.11 family of protocols. By taking advantage of various diversity techniques, our scheme effectively increases the SINR of received signals, which helps improve the accuracy of synchronization as well as mitigate the impact of asynchronicity in a distributed scenario. As in [13], random access codes are used to mask control signals which are transmitted simultaneously from selected nodes, and used for transmission coordination and channel estimation.

RTS In RTS transmission phase, nodes that determine themselves to be transmitter nodes (using algorithm in Section 5.1.1) broadcast RTSs. For a transmitter node n_j, the RTS message contains the number of streams it plans to transmit N_j^0, its node ID and the IDs of the destination nodes. The preamble of a packet is used as the training sequence (without incurring extra overhead for adding in pilot signal) for channel estimation purpose. The preamble of an RTS message is transmitted rotationally from each antenna so the full channel condition matrix can be estimated at receiver nodes. RTS messages sent from different transmitters are masked by different ID code to allow a receiver to differentiate the messages. As the number of antennas is generally small and only the preamble of the RTS message is transmitted through all antennas, the total transmission delay for channel estimation purpose is small. The full knowledge of the channel as a result of the estimation, however, could enable simultaneous transmission of multiple spatial streams and bring in multi-fold capacity gain [13] and thus delay reduction.

CTS The RTSs are received at receiver nodes, where channel matrices are estimated by extracting the preambles. A receiver node n_m also estimates the number of streams it may receive $N_m^{rec} = \sum_{j \in \mathscr{V}_m, x_j=1} N_j^0$. Constrained by its degree of freedom, n_m can decode at most N_m^{dec} streams simultaneously. If n_m receives multiple RTSs (from the source and/or candidate relay nodes) on the transmission of p_i in current TD and is the target receiver of p_i, it then selects the node $r(p_i)$ which has the best channel condition between $r(p_i)$ and n_m to forward the packet. Based on the decoding capability and the signal strength received, n_m estimates the interference plus noise level (SINR) for candidate transmission nodes. Finally, n_m broadcast a CTS message including SINR, N_m^{rec}, N_m^{dec} and $r(p_i)$. Note that CTS message is also masked by ID code and the preamble is transmitted rotationally from each antenna of n_m for transmitter nodes to estimate the full channel condition matrix.

DATA In the DATA slot, a sender first determines the number of streams it is allowed to transmit using the algorithm in Section 5.1.2, based on the information received from CTSs sent by neighboring receivers. It then estimates the transmission rate from each antenna based on the estimated channel condition and interference at a destined receiver, and transmits the packets from the antennas selected using the maximum weight matching algorithm in Section 5.2. A receiver node then differentiates all streams it receives and extracts the data packets targeted for it. Instead of discarding packets transmitted through interference streams, a receiver buffers an overheard packet if it is within the transmission range of the packet destination for potential relay transmission.

ACK-I Receiver nodes broadcast ACKs about those successfully received packets, which include the original sources of the packets. These messages are received by nodes that are not receivers in the current TD.

ACK-II If a relayed packet is received successfully, the source node as well as all the potential relay nodes should remove it from their buffers in order to avoid redundant transmissions. Some of these nodes may not be able to receive the ACKs as they are also in transmitting states during the transmission of ACKs. After the transmission of ACK-I, ACK-II is rebroadcast by non-receiver nodes in the current TD. With the transmission of ACKs in consecutive phases, it not only ensures all candidate relay nodes to learn the packet transmission status, but also allows a relay node to inform the original packet sender the successful transmission of the relay packet. This is important. As the channel condition between the source and the destination is not good, the loss of ACK-I message sent from the destination could lead to continuous retransmissions at the source and waste more wireless resources.

7. Performance Evaluation

In this section, we evaluate the performance of our proposed algorithms through simulations. We consider an ad hoc network with random topology where nodes are distributed uniformly over a $1250m \times 1250m$ area. Each node has a reference transmission range of $250m$ as in a standard 802.11 wireless network. Both path loss and independent Rayleigh fading are incorporated for a wireless link between an antenna pair. For each node, the number of incoming data packets is Poisson distributed with a given mean value λ and the destination of each packet is chosen at random. A simulation result is obtained by averaging over ten runs of simulations with different seeds.

The two-phase scheduling algorithm proposed in Section 5 is implemented based on the MAC protocol described in Section 6. The cooperative relayed spatial multiplexing schemes proposed in this paper are named as CRSM-C or CRSM-D respectively, depending on whether a centralized scheme or a distributed scheme is used for the determination of transmitter nodes and the number of transmission streams. Correspondingly, the opportunistic and cooperative spatial multiplexing (OCSM) scheme proposed in [13] which does not involve a relay transmission is also implemented for comparison purpose. The metrics we use are normalized throughput and delay. Throughput is the total effective data rate of the network averaged over the number of transmission durations, which takes into account the impact of control overhead. Delay time is defined as the number of transmission durations a packet waits in the queue before it is removed from the MAC queue. The transmission delay includes the time for transmission of control packets. We investigate the impact on network

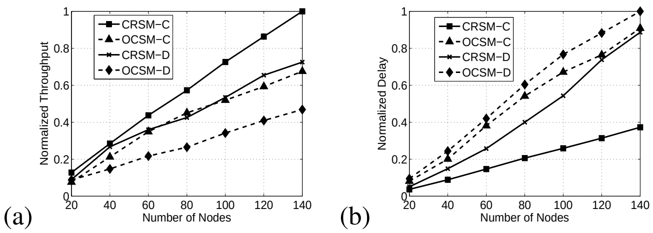

Figure 2. Impact of node density: (a) data rate; (b) delay.

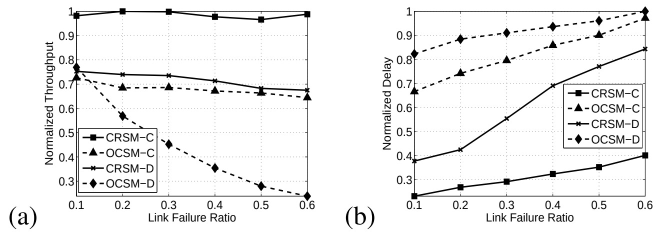

Figure 3. Impact of link failure ratio: (a) data rate; (b) delay.

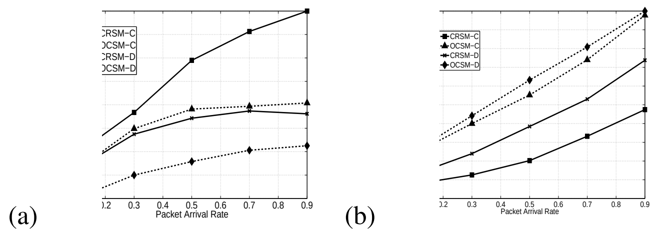

Figure 4. Impact of packet arrival rate: (a) data rate; (b) delay.

performance due to four factors, namely node density, link failure ratio, packet arrival rate and retransmission threshold. The retransmission threshold defined in Section 6.1 is in the unit of TD, and a packet is dropped from both the source queue and queues of candidate relay nodes when the time lasted from the initial packet transmission exceeds the threshold. If not otherwise specified, the number of nodes in the network is 100, the link failure ratio is 0.3, the average packet arrival rate λ is 0.5 and the retransmission threshold is 8.

7.1. Impact of Node Density

The impact of node density is shown in Fig. 2. Increased node density leads to heavier traffic and also provides more links among nodes in a network. In case of severe links, the two CRSM schemes have a higher possibility of finding candidate relay nodes to assist in transmission by taking advantage of the improved connectivity. In Fig. 2 (a), CRSM-D is observed to improve the throughput up to 53% compared to OCSM-D. Effective scheduling of packets with relay also reduces the queuing delay as seen in Fig. 2 (b).

7.2. Impact of Link Failure Ratio

A link is considered to be failed if a packet transmitted through it can not be received successfully by its receiver. Link failure can be a result of path loss, deep fading of channels, mobility of nodes, etc. We use link failure ratio (LFR) to model the percentage of failed links over all links in the network. The two CRSM schemes are shown to have a robust performance under different link failure ratios, as in Fig. 3. In Fig. 3 (a), while the throughput of OCSM-D degrades tremendously with increasing LFR, only a slight throughput degradation is observed with both CRSM schemes. The throughput of CRSM-D is three times that of OCSM-D when a frequent link breakage occurs at $LPF = 0.6$, and the delay reduction is up to 50%. A higher link breakage ratio would lead to increased delay. The significant performance improvement demonstrates the effectiveness of using relay in MIMO transmissions to improve reliability in a harsh transmission environment.

7.3. Impact of Packet Arrival Rate

The mean packet arrival rate λ captures the traffic load in a network. By adaptively using cooperative relay transmissions, high rate links are more efficiently utilized to schedule heavier traffic load. In Fig. 4 (a), even with the heaviest traffic load, CRSM-D still achieves 35.7% higher throughput than OCSM-D. Although higher traffic increases queuing delay of packets due to limited network capacity, the normalized delay of CRSM-D scheme is about 30% lower than that of OCSM-D. This demonstrates that relay can be used effectively in networks with heavy traffic load to improve performance.

7.4. Impact of Retransmission Threshold

Retransmission is a common strategy used to deal with temporary transmission failure. The performances of CRSM and OCSM are further compared in Fig. 5 under different values of the retransmission threshold F, as introduced in Section 6. In CRSM schemes, packets experienced direct transmission failure can be forwarded through relay links which may have better link conditions than the direct link. With increased value of F, both CRSM schemes keep a nearly constant throughput values, while OCSM-D undergoes significant throughput reduction. Even though more retransmissions help increase the probability of successful packet reception, transmissions over poor links for a longer period of time would consume more network resources. On the contrary, both CRSM schemes actually take advantage of a large F to conduct relay transmissions through adaptive scheduling. The delays of two OCSM schemes and CRSM-D scheme all increase with F with the increase of time to keep the packets in buffers, while CRSM-D remains to have much lower delay than OCSM-D under all values of F.

8. Conclusion

Ad hoc networks are popularly used in military and emergency rescue environments. In addition, there are increasing interests in applying ad hoc networks to connect

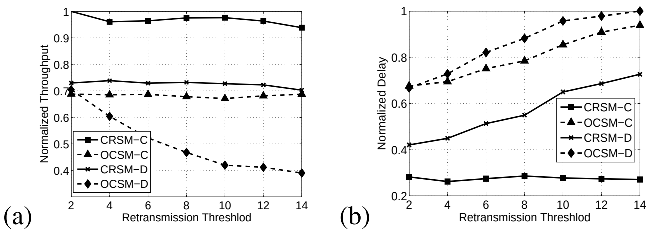

Figure 5. Impact of retransmission threshold: (a) data rate; (b) delay.

various wireless devices to enable more powerful wireless applications and mobile computing capabilities. All these applications require higher network throughput and reliability. In this work, we design scheduling algorithms and MAC protocol to enable cooperative relay transmission in MIMO-based ad hoc networks, in order to jointly exploit the cooperative multiplexing gain and cooperative diversity gain to achieve overall higher data rate and lower delay under harsh channel conditions. We formulate the problem of packet scheduling with cooperative relay in MIMO ad hoc networks as an integer programming problem, and propose both centralized and distributed solutions to support relay transmissions. We also design an effective MAC protocol to facilitate the implementation of the distributed scheduling algorithm. Through extensive simulations, our scheme is shown to outperform the reference MIMO scheme which does not use relay, with significantly higher throughput and reduced average delay. This demonstrates the importance of incorporating relay transmissions in MIMO-based ad hoc networks and the effectiveness of the proposed algorithm in enabling concurrent MIMO and relay transmissions.

References

[1] G. J. Foschini. Layered space-time architecture for wireless communication in fading environments when using multi-element antennas. *Bell Labs Tech. J.*, (2):41–59, 1996.

[2] S. M. Alamouti. A simple transmit diversity technique for wireless communications. *IEEE J. Select. Areas Commun.*, pages 1451–1458, October 1998.

[3] V. Tarokh, N. Seshardi, and A. Calderbank. Space-time codes for high data rate wireless communication: Performance criteria and code construction. *IEEE Trans. Inform. Theory*, (2):744–765, March 1998.

[4] IEEE Strandards. *IEEE Standard 802.11n*.

[5] M. Hu and J. Zhang. MIMO ad hoc networks: Medium access control, saturation throughput, and optimal hop distance. *Journal of Communications and Networks, Special Issue on Mobile Ad Hoc Networks*, pages 317–330, 2004.

[6] R. Bhatia and L. Li. Throughput optimization of wireless mesh networks with MIMO links. In *Proc. IEEE INFOCOM 2007*, May 2007.

[7] M. Levorato, S. Tomasin, P. Casari, and M. Zorzi. Analysis of spatial multiplexing for cross-layer design of MIMO ad hoc networks. In *Proc. IEEE VTC-2006 Spring*, pages 1146–1150, May 2006.

[8] M. Park, R. Heath, and S. Nettles. Improving throughput and fairness of MIMO ad hoc networks using antenna selection diversity. In *Proc. IEEE Globecom 2004*, November 2004.

[9] K. Sundaresan, R. Sivakumar, M. Ingram, and T-Y. Chang. A fair medium access control protocol for ad-hoc networks with MIMO links. In *Proc. IEEE INFOCOM 2004*, June 2004.

[10] G. Jakllari, S. V. Krishnamurthy, M. Faloutsos, P. V. Krishnamurthy, and O.A Ercetin. Framework for distributed spatio-temporal communications in mobile ad hoc networks. In *Proc. IEEE INFOCOM 2006*, pages 1–13, April 2006.

[11] B. Hamdaoui and P. Ramanathan. A cross-layer admission control framework for wireless ad-hoc networks using multiple antennas. *IEEE Trans. on Wireless Communications*, November 2007.

[12] K. Sundaresan and R. Sivakumar. Cooperating with smartness: Using heterogeneous smart antennas in ad-hoc networks. In *Proc. INFOCOM 2007*, pages 303–311, May 2007.

[13] S. Chu and X. Wang. Opportunistic and cooperative spatial multiplexing in MIMO ad hoc networks. In *Proc. ACM MobiHoc 2008*, pages 63–72, May 2008.

[14] A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversity part I and part II. *IEEE Trans. Commun.*, (11):1927–1948, November 2003.

[15] 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*, (12):3062–3080, December 2004.

[16] S. Cui, A. J. Goldsmith, and A. Bahai. Energy-efficiency of MIMO and cooperative MIMO in sensor networks. *IEEE J. Select. Areas of Commun.*, (6):1089–1098, August 2004.

[17] Y. Shi, S. Sharma, Y. T. Hou, and S. Kompella. Optimal relay assignment for cooperative communications. In *Proc. ACM Mobihoc 2008*, May 2008.

[18] M. Yuksel and E. Erkip. Multiple-antenna cooperative wireless systems: A diversitycmultiplexing tradeoff perspective. *IEEE Trans. Inform. Theory*, Octorber 2007.

[19] M. Levorato, S. Tomasin, and M. Zorzi. Cooperative spatial ultiplexing for ad hoc networks with hybrid ARQ: system design and performance analysis. *IEEE Trans. Commun.*, (9):1545–1555, September 2008.

[20] Y. Pan and S. Aissa. Performance analysis of selective space-time coding and selection diversity under perfect and imperfect CSI. In *Proc. PIMRC 2005*, pages 2371–2375, September 2005.

[21] D. Tse and P. Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, May 2005.

[22] Mischa Schwartz. *Mobile Wireless Communications*. Cambridge University Press, 2005.

[23] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network flows: theory, algorithms and applicationss*. Prentice Hall, 1993.