Adaptive Scheduling in MIMO-based Heterogeneous Ad hoc Networks

Shan Chu, Xin Wang *Member, IEEE,* and Yuanyuan Yang *Fellow, IEEE.*

Abstract—The demands for data rate and transmission reliability constantly increase with the explosive use of wireless devices and the advancement of mobile computing techniques. Multiple-input and multiple-output (MIMO) technique is considered as one of the most promising wireless technologies that can significantly improve transmission capacity and reliability. Many emerging mobile wireless applications require peer-to-peer transmissions over an ad hoc network, where the nodes often have different number of antennas, and the channel condition and network topology vary over time. It is important and challenging to develop efficient schemes to coordinate transmission resource sharing among a heterogeneous group of nodes over an infrastructure-free mobile ad hoc network. In this work, we propose a holistic scheduling algorithm that can adaptively select different transmission strategies based on the node types and channel conditions to effectively relieve the bottleneck effect caused by nodes with smaller antenna arrays, and avoid the transmission failure due to the violation of lower degree of freedom constraint resulted from the channel dependency. The algorithm also takes advantage of channel information to opportunistically schedule cooperative spatial multiplexed transmissions between nodes and provide special transmission support for higher priority nodes with weak channels, so that the data rate of the network can be maximized while user transmission quality requirement is supported. The performance of our algorithm is studied through extensive simulations and the results demonstrate that our algorithm is very effective in handling node heterogeneity and channel constraint, and can significantly increase the throughput while reducing the transmission delay.

Index Terms—Mobile computing, ad hoc networks, scheduling, distributed, cross-layer design, MIMO.

1 Introduction

There are increasing interests and use of mobile ad hoc networks with the proliferation of mobile, network-enabled wireless devices, and the fast progress of computing techniques and wireless networking techniques. As the number, CPU power and storage space of wireless devices continue to grow, there is a significant increase in data transmission demand to support data intensive mobile computing and applications, such as multimedia streaming, gaming, as well as transmission of a large amount of monitoring data. To meet the high data rate requirements, more and more wireless devices are equipped with multiple antennas. With multiple antennas at the transmitter and/or receiver, a MIMO (multiple-input-multiple-output) system takes advantage of *multiplexing* to simultaneously transmit multiple data streams to increase the wireless data rate and *diversity* to optimally combine signals from different transmission streams to increase the transmission reliability and range. The benefits of MIMO lead many to believe it is the most promising technique of emerging wireless technologies. MIMO is prominently regarded as a technology of choice for next generation wireless systems such as IEEE 802.16, IEEE 802.11n, and the third and fourth generation cellular systems. It is also being considered for supporting peer to peer mobile applications over an infrastructure free ad-hoc network.

Although MIMO techniques have been widely studied in a more centralized and infrastructure-based cellular system, there are very limited work and big challenges in extending MIMO technique into a fully distributed system over an infrastructure-free wireless ad hoc network. Different from an infrastructure-based system, it is difficult for nodes to coordinate in channel evaluations and transmissions in a distributed manner. The fast variation of channel condition and network topology, the inconsistency in node density as well as the different traffic demands and service requirements of nodes lead to more open challenges in coordinating distributed node transmissions.

Moreover, in a mobile computing environment, the network could be *heterogeneous*, which incurs additional challenges to MIMO MAC design. First, network nodes may be equipped with different number of antennas. The existence of nodes with smaller antenna array sizes may lead to significant network performance reduction. Either the concurrent number of transmissions in a neighborhood needs to be limited in order to meet the decoding constraint of receivers equipped with a lower number of antennas, or the lower-antenna nodes will be significantly interfered by neighboring nodes transmitting a larger number of streams at the same time. Second, the transmission environment could be heterogeneous, with channel conditions different between each node pair and varying over time, leading to the variation of the simultaneous streams allowed between a node pair. These two factors jointly determine the number of orthogonal channels (i.e., *degree of freedom*) an environment allows. It is critical for transmitter nodes to be aware of the allowed degree of freedom of a link for the correct decoding at receiver nodes.

Recently, there have been some efforts in developing algorithms and protocols for applying MIMO techniques to ad hoc networks [1]–[10], however, it is far from trivial to extend the solution in homogeneous cases to heterogeneous cases especially in a distributed system. To the best of our knowledge, there has not been any effort to specifically alleviate the transmission limit thus performance degradation due to the network heterogeneity in a distributed, peer to peer, ad hoc transmission environment.

To enable more powerful mobile computing and applications in a practical system, the objective of this work is

• *Shan Chu is with Motorola Solutions, One Motorola Plaza, Holtville, NY 11742. E-mail: shan_chu@hotmail.com. Xin Wang and Yuanyuan Yang are with the Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY 11794. E-mail: {xwang,yang}@ece.sunysb.edu.*

to design a holistic framework for scheduling to adaptively coordinate sharing of transmission resources among heterogeneous nodes in a varying physical operational scenario. The main contributions of this paper are as follows.

- Our scheduling algorithm concurrently considers antenna array size, channel condition, traffic demand and multiuser diversity. In each transmission duration, the algorithm opportunistically schedules the nodes to transmit and determines the set of antennas to use at a selected node in a distributed manner, exploiting multiuser diversity and antenna selection diversity to significantly improve the transmission throughput and reliability. Through priority-aware scheduling, our algorithms also support service differentiation while reducing the transmission delay and ensuring the fairness among nodes. As interference alignment and transmitter side interference cancelation techniques bring too much computational overhead and are difficult to apply in a mesh network scenario, we only consider receiver-side interference cancelation in this work.
- Our algorithms specifically alleviate the constraints caused by node heterogeneity and the lower-rank channel, by adaptively selecting different transmission strategies based on both the antenna array sizes of nodes in a neighborhood and the degree-of-freedom the transmission channel allows.
- We mathematically formulate the problem to maximize the weighted network throughput, and propose a centralized scheduling algorithm as a performance benchmark. Different from the literature work [3]–[5], [11], [12] which are based on the simple antenna model, our formulation takes into account the different transmission rates between different nodes and antenna pairs and the constraint on the degree of freedom due to the channel condition between a node pair. Our scheduling algorithm thus can select nodes and antennas with better transmission rate in a time slot, taking advantage of multi-user diversity and antenna selection diversity for a higher throughput. The algorithm also considers the transmission priority and balances the network load, while avoiding transmission failure by not overloading a lower-rank channel with more streams.
- We perform extensive simulation to investigate the impact of various factors on the performance of MIMO scheduling, and to evaluate the effectiveness of our proposed algorithm. Specifically, to examine the impact of channel heterogeneity, we studied the impact on the spatial channel both due to antenna array size and line-of-sight component. By considering more realistic channel conditions and network heterogeneity, our results are expected to provide some insight for the implementation of the algorithms in practical networks. Our simulations consider both the static network topology and a dynamic network topology with mobility modeled by the random waypoint model. However the model dynamics affect only path loss, and do not change our assumption of quasi-static Rician fading during a transmission duration.

The rest of the paper is organized as follows. We present the related work in Section 2. Section 3 discusses the background information including MIMO technologies and their application in heterogeneous networks. In Section 4, the system model is defined and the problem is mathematically formulated, followed by Section 5, where a centralized algorithm is proposed to solve it. Section 6 presents the adaptive distributed scheduling algorithm and the protocol to implement the algorithm. Simulation results are provided in Section 7 and the paper is concluded in Section 8.

2 Related Work

Over the past several years, the application of MIMO technology in networks has undergone a fast development. Earlier studies have been performed to develop scheduling schemes to select the best users to transmit based on certain criteria in a multiuser MIMO-based cellular network. An overview of scheduling algorithms in MIMO-based fourth-generation wireless systems is given in [13].

In recent years, many efforts have been made to support MIMO transmission in local area wireless networks. In [1], spatial diversity (e.g. space time coding (STC)) is explored to combat fading and achieve robustness. SPACE-MAC, proposed in [2], enables denser spatial reuse patterns with the aid of transmitter and receiver beamforming. In [14], the design and implementation of a cross-layer system is presented to enable spatial multiplexed transmissions from multiple devices to the access point in the uplink direction in a wireless LAN environment, while coordinating node transmissions in a multi-hop meshed network is much more challenging, and the challenge increases if the network is heterogeneous. Recently, there have also been some efforts in applying MIMO with interference cancellation and alignment into wireless networks through testbed study [15]–[17] for MIMO application in wireless LANs. However, wireless LANs are different from wireless mesh and ad hoc networks in the network scale. The algorithm and framework for wireless LANs cannot be simply applied to multi-hop wireless networks, especially in the challenging distributed case. Specifically, interference alignment require significantly more complicated coordination and extra information of channel conditions in a multi-hop ad hoc network scenario, which makes it difficult to apply in a practical network. Moreover, although interference alignment is considered to provide the maximum degree of freedom theoretically, it does not guarantee to achieve the maximum throughput. Therefore, it is out of the scope of this work.

Due to the difficulty of modeling the benefits and constraints of MIMO transmissions in ad hoc networks, only a limited number of efforts focus on the network performance from the optimization perspective. A centralized algorithm is presented in [3] to solve the joint routing, scheduling and stream control problem subject to the fairness constraint in mesh networks with MIMO links. In [4], the authors characterize the radio and interference constraints in multi-hop wireless MIMO networks and formulate a multi-hop joint routing and MAC problem to study the maximum achievable throughput subject to these constraints. The problem of jointly optimizing power and bandwidth allocation at each node and multi-hop/multi-path routing in a MIMO-based ad hoc network is studied in [5], and a solution procedure is developed to solve this cross-layer optimization problem. In [11], [12], the authors study the link layer model for multi-hop MIMO networks based on accurate accounting of DoF consumption and a node-level ordering scheme was proposed to identify the role of each node in performing interference cancellation. In [18], optimal stream scheduling for MIMO links is studied for a single collision domain, and it was shown that optimum throughput is achieved when the task of interference cancellation is shared equally between every transmitter and every receiver. Although these efforts are important, the aforementioned work [3], [4], [11], [12] have assumed simplified physical model and overlooked the impact of channel condition by assuming streams have homogeneous data rate. In fact, these simplifications may not only significantly compromise the network performance, but also make the optimal model formulated far from the practical network condition. In addition, none of them provides a feasible solution to efficiently coordinate node transmissions in a practical distributed scenario. The features and performance of a few antenna techniques are presented in [6], however there is no design to enable the selection of a specific antenna technique, which is the major challenge in MAC design.

A number of distributed schemes have also been proposed for MIMO MAC designs. In [7], the authors discuss key considerations for MIMO MAC design, and develop a centralized algorithm and a distributed algorithm to improve the transmission fairness. Based on CSMA/CA for control signal exchanges, it is hard for the algorithm to support cooperative transmissions. In [8], spatial multiplexing with antenna subset selection for data packet transmission is proposed, based on nodes with two antennas and a simple network topology. In [9], each transmitter or receiver greedily accommodates the number of data and interference streams up to a pre-determined maximum number. However, it is possible for different receivers to make conflicting decisions on the transmission requests to accept, wasting the decoding capability or exceeding the decoding limit of the receivers . In addition to issues associated with each scheme discussed above, most MIMO MAC schemes implicitly assume that the channel condition is known. In practice, coordinating channel measurement itself is a big challenge in presence of a group of competing nodes, especially in a distributed and a dynamic ad hoc network. In existing work, the number of antennas or the pre-determined decoding limit is often used as the constraint of transmission and receiving without considering the actual physical channel variation thus the simultaneous streams allowed by a channel, which may result in the transmission failure. In [10], a cooperative multiplexing scheme is proposed, however, it does not consider the heterogeneity of antenna arrays in ad hoc networks. We have made an effort to provide an adaptive and distributed solution considering the heterogeneity of antenna array sizes of network nodes in [19]. In this paper, we further investigate the impact of channel condition on the degree of freedom of MIMO channels. We remodel the problem to more accurately capture the transmission constraints due to both the number of antennas and channel conditions. We also modify the distributed algorithms and perform more extensive simulations to demonstrate the functionality of the proposed distributed algorithms. In addition, we propose a centralized solution with a proved approximation ratio to serve as the benchmark of the distributed algorithm.

3 Background and Motivation

As mentioned in Section 1, with multiple antennas at the transmitter and/or receiver, multiple data streams may be transmitted between a transmission node pair, which is called *spatial multiplexing*. At the receiver, each antenna receives a superposition of all of the transmitted data streams. In a rich scattering environment where the transmission channels for different stream are differentiable and independent, i.e. orthogonal, an intended receiver node can separate and decode its received data streams based on their unique spatial signatures. This multiplexing gain can provide a linear increase (in the number of antenna elements) in the asymptotic link capacity. With multiple transmission paths, the transmission quality could be very different. Instead of sending different data through each transmitting antenna, *spatial diversity* may be exploited to improve transmission reliability. There are different types of diversity techniques. Without channel information, dependent streams can be transmitted on different antenna elements over multiple time slots and improve transmission quality through *space time coding*. 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]. As a more powerful yet more sophisticated scheme, data streams can be properly coded according to the channel information and sent through different transmit antennas, i.e., through *precoding*, to achieve the maximum throughput at the receiver.

In this section, we first present the problems due to the limitation of channel degree of freedom and heterogeneous number of network nodes. We then introduce the potential strategies to address the issues, and the tradeoff between different strategies.

In MIMO communications, the spatial channels between two neighboring nodes n_i and n_k which have N_i^{ant} and N_k^{ant} antenna elements respectively can be represented as a $N_k^{ant} \times N_i^{ant}$ matrix:

$$\mathbf{H}_{ki} = \begin{pmatrix} h_{11} & h_{12} & \dots & h_{1N_i^{ant}} \\ h_{21} & h_{22} & \dots & h_{2N_i^{ant}} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_k^{ant}1} & h_{N_k^{ant}2} & \dots & h_{N_k^{ant}N_i^{ant}} \end{pmatrix}. \quad (1)$$

The (p,q)-th entry of $\mathbf{H}_{ki}$, h_{pq}, is the spatial channel coefficient between the p-th antenna of node n_k and q-th antenna of node n_i. Each h_{pq} can generally be represented as [21]:

$$h_{pq} = \sqrt{\frac{\kappa}{\kappa+1}}\sigma_l e^{j\theta} + \sqrt{\frac{1}{\kappa+1}}\mathcal{CN}(0,\sigma_l^2), \quad (2)$$

where the first term denotes the line-of-sight (LOS) component with a uniform phase θ, and the second term corresponds to the aggregation of reflected and scattered paths, usually modeled as a circular symmetric random variable with variance σ_l. The parameter κ is called K-factor, which is the ratio of the energy in the LOS path to the energy in the scattered paths. When the LOS component is very weak, i.e. the propagation medium is rich scattering, the channel can be well modeled by Rayleigh fading. When the LOS component between transmitter and receiver is strong and/or there exist fixed scatters/signal reflectors in addition to random main scatters, Rician fading conditions hold and a higher correlation is observed between the elements of $\mathbf{H}_{ki}$.

The *degree-of-freedom* of a MIMO channel is an important metric to describe the dimension of space that the transmitted signals can be projected onto (so the receiver can differentiate the signals), and the number of streams allowed to simultaneously transmit between a pair of nodes. The

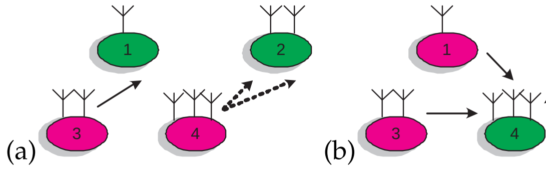

Fig. 1. Illustration of a heterogeneous MIMO network.

degree-of-freedom is defined as the *rank* of the channel matrix $\mathbf{H}_{ki}$, or equivalently the number of non-zero eigenvalues of $\mathbf{H}_{ki}$. From (1), it is obvious that the degree-of-freedom of the channel between n_i and n_k depends on the number of antennas at nodes n_i and n_k, and the linear independency of the matrix which depends on the scattering conditions between n_i and n_k.

In Fig. 1, the four nodes, each equipped with an antenna array, are in the transmission range of each other. In (a), if node 1 is a selected receiver in a time slot, in order to ensure its correct decoding only one stream targeted to 1 is allowed to transmit in its neighborhood. Moreover, when both nodes 1 and 2 are selected as receivers, even though node 4 would be able to transmit up to 2 streams to node 2 (as shown in dashed lines), if simply scheduling the transmissions based on the minimum number of streams allowed in a neighborhood [10], only one stream is allowed to be transmitted around node 2 at a transmission time (e.g., either transmitting from node 3 or from node 4). That is, without differentiating node types, the maximum number of streams allowed to transmit at any time slot is constrained by the candidate receiver which has the smallest array. On the other hand, if every receiver simply considers its own decoding constraint [9], a higher number of transmissions could lead to serious interference and potential decoding failure at nodes with a lower number of antennas. In addition, when the channel between node 4 and 2 can only support one transmission, i.e. the degree-of-freedom is 1, but two streams are transmitted, the streams cannot be decoded at the receiver. The examples indicate that it would lead to either significant throughput reduction in order to not interfere with a node with lower number of antennas or transmission failure if the node heterogeneity and channel rank constraint are not considered in the MAC design. Additional issues will arise if some of the channels are weak, and cannot support good quality transmission.

These practical problems indicate that effective scheduling algorithms need to be designed to alleviate the bottleneck effect and to provide good system performance under any transmission environment. A few strategies may help. First, when the receiver has multiple antennas, the constraints to transmissions due to the lower degree-of-freedom between node pairs may be mitigated with the formulation of cooperative virtual MIMO array. In Figure 1(b), nodes 1, 2 and 3 can transmit concurrently to node 4 and exploit multiplexing gain to improve the throughput. Second, additional capacity gain can be achieved with the exploration of multi-user diversity and antenna selection diversity, in which case, the transmitter nodes and the antenna to use from a node are opportunistically selected based on the channel conditions between different nodes and antennas. Third, when the receiver has very few antennas (Node 1 in Fig. 1 (a)), its transmitter could employ precoding to optimally weight the transmissions from multiple antennas to improve the data rate.

As shown in the example, opportunistic multiplexing transmission generally allow multiple nodes to 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. The many-to-many nature of the transmission makes precoding difficult to be applied across multiple transmitters and receivers, as the calculation of the precoding weights involves multiple channel matrices and is much more complicated than the one-to-many case in cellular network. Therefore, precoding it is not used simultaneously with the cooperative multiplexing. In Fig. 2, a simple experiment is performed to compare the data rates achieved by opportunistic transmitter precoding (OTPC) and opportunistic and cooperative spatial multiplexing (OCSM) [10] under a topology where two transmitter nodes are around one receiver node with i.i.d faded channels. The performances of the two are compared with the variation of the number of antennas at each node. When the receiver antenna array size is small, transmitter precoding is seen to outperform multi-user multiplexing as power gain is more significant. However, with more receiving antennas, cooperative multiplexing starts to outperform precoding. From this simple example, we can see that it is important to select an appropriate transmission strategy according to specific constraints, in order to achieve optimum possible performance. Instead of transmitting the same signal from multiple antennas with appropriate weighting to increase the rate of one stream as done in conventional beam-forming scheme, in this work each data packet is transmitted only through one selected stream and selected streams from all candidate antenna pairs form many-to-many cooperative MIMO transmissions to improve the total network capacity. Therefore, precoding is only assumed to weight the transmissions when multiple streams are selected to transmit between a node pair.

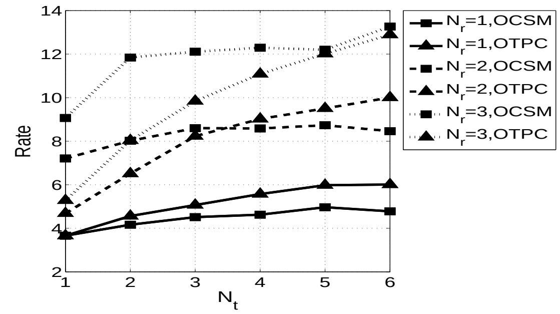

Fig. 2. Comparison of multiplexing and transmitter precoding with varied transmitter/receiver antenna array size.

4 System Model

We consider channel resource allocation among an ad hoc network of nodes which have different number of antenna elements and experience different channel conditions. For a group of nodes that share the transmission resource, one node pair is often scheduled to transmit at a time in the traditional MIMO schemes. However, the chance of having multiple strong spatial paths between a node pair is small, which limits the transmission rate. Instead, our scheduling schemes support many-to-many transmissions between nodes using virtual MIMO arrays, and take advantage of multi-user diversity and antenna selection diversity to significantly improve the transmission reliability and throughput. Specifically, to address the challenges due to the network heterogeneity, our algorithms *adaptively and flexibly* schedule node transmissions using different MIMO techniques, including spatial multiplexing, selection diversity, and precoding, based on the node constraints and channel conditions.

For the convenience of presentation, in this section, we first introduce some notations used in this paper, and then formulate the problem mathematically and prove its NP-hardness.

4.1 Stream and Stream Characteristics

A stream is defined to be an *independent* flow of signals transmitted from a transmit antenna to a target node and identified by a triplet $(I^{tx}, I^{rc}, I^{ant})$, where I^{tx}/I^{rc} is the index of the transmitter/receiver node, and I^{ant} is the index of the antenna that involves in the transmissions of the stream. With the exploitation of selection diversity, the antennas with the strongest channel conditions among the candidate ones are selected to transmit the data streams. For a transmitter node with several streams selected, if the streams target for the same receiver, precoding is performed among the selected transmitting antennas with the power optimally allocated to achieve the maximum data rate between a node pair; otherwise, the power is evenly distributed over the selected antennas for streams targeting for different receivers for the processing simplicity.

In order for receiver nodes to decode data streams and suppress interference streams concurrently, the number of streams transmitted or received at a node is subject to certain constraint. Due to the broadcast nature of wireless channels, streams are categorized as *data streams* and *interference streams*. A data stream from node n_i to node n_k is received by n_i's neighboring node n_j as an interference stream. Denote the degree-of-freedom of the channel between n_i and n_k as $DoF(i,k)$, it is clear that n_k can differentiate streams from n_i only if the number of streams is no more than $DoF(i,k)$, which depend on both the antenna numbers of n_i and n_k and the correlation level of the channel between the two nodes. Denote the set of all active receiving nodes (i.e., the target receivers of some transmitter nodes) around node n_i's transmission range as R_i^{active}, as the *transmitting constraint*, the number of transmitting streams from n_i should be no larger than $\mathcal{N}_i^{tx} = \min_{k \in R_i^{active}} DoF(i,k)$. Note that transmitter side interference cancellation is out of our scope here. Similarly, to avoid erroneous decoding at a receiver node n_k, the number of simultaneous received streams N_k^{rc} (including both data streams and interference streams) should be limited. With use of virtual MIMO array, the size of antenna array N_k^{ant} generally provides the metric of spatial resolution at a receiver n_k, and hence the total received streams should not exceed the *receiving constraint* $\mathcal{N}_k^{rc} = N_k^{ant}$.

The characteristics of a stream are captured by two parameters, *stream priority* $\mathscr{P}(s)$ and *stream capacity* $\mathscr{C}(s)$. The stream priority depends on the service type and queuing delay of the data packet to be sent with the stream. The value of $\mathscr{P}(s)$ is initially set to the service priority of the associated packet, and increases as the queuing time of the packet increases. The stream capacity describes the maximum achievable rate of a stream transmission, which depends on receiving signal to interference and noise ratio of the stream, and can generally be represented as follows:

$$\mathscr{C}(s) = \log\left(1 + P_s \mathbf{h}_s^* \left(N_0 \mathbf{I}_{N_{d(s)}^{ant}} + \sum_{q \in I(s)} P_q \mathbf{h}_q \mathbf{h}_q^*\right)^{-1} \mathbf{h}_s\right), \tag{3}$$

where P_s is the transmission power of the stream s, $\mathbf{h}_s$ is the channel vector from the selected transmitter antenna and the receiver of stream s, N_0 is the noise variance, $N_{d(s)}^{ant}$ is the number of antennas at the receiver of stream s and $I(s)$ is the set of streams that interfere the receiving of stream s at the receiver. The value of $\mathscr{C}(s)$ can be estimated at a transmitter based on the estimated channel condition and interference level during the scheduling.

4.2 Types of Nodes and Slots

Our algorithm is TDMA-based, in which the time domain is divided into transmission durations (TD). A TD consists of several time slots and covers one round of control signal exchange and fixed-size 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. A link between a transmitter-receiver pair is half-duplex, so that a node can either transmit or receive but not at the same time.

Denote the set of nodes in the transmission range of node n_i as $\mathscr{V}_i^A$, the receiving constraint of node n_k as $\mathcal{N}_k^{rc}$. Since a node with a higher value of $\mathcal{N}_k^{rc}$ can generally decode more streams, we use $\mathcal{N}_k^{rc}$ as a metric for measuring the receiving capability of n_k. The average receiving capability of nodes in $\mathscr{V}_i^A$ is then represented as $\bar{\mathcal{N}}_i^{rc} = \frac{1}{|\mathscr{V}_i^A|}\sum_{k \in \mathscr{V}_i^A} \mathcal{N}_k^{rc}$. Compared with $\bar{\mathcal{N}}_i^{rc}$, if $\mathcal{N}_k^{rc} \geq \bar{\mathcal{N}}_i^{rc}$, node n_i considers n_k as a *rich* node as it has relatively higher receiving capability among the neighboring nodes of n_i; otherwise, n_k is considered as a *poor* node and could potentially become a receiving bottleneck in the neighborhood. Note that when all nodes have the same number of antennas, the network contains only rich nodes, and it is degenerated to the homogeneous network case.

As discussed in Section 3, the limited decoding capability of a poor receiver constrains the maximum number of streams (including both data streams and interference streams) allowable in its neighborhood. To reduce the constraint, we divide the transmission slots into *P-slots* and *R-slots* and assume different transmission strategies towards poor nodes and rich nodes respectively. In a P-slot, the number of concurrent transmission streams is limited by the receiving constraint of the targeted poor node, and transmitter precoding may be utilized to optimize the link rate. In an R-slot, as only rich nodes serve as the receivers, multiuser spatial multiplexed transmissions are opportunistically scheduled for a higher throughput.

4.3 Problem Formulation

In a TDMA-based MIMO ad hoc network, packets are generated constantly. It is thus practical to schedule the transmission of packets in each transmission duration (TD) with the purpose of optimizing temporary network performance. Suppose there is a set of nodes $N = \{n_1, n_2, \ldots, n_{N_n}\}$ in the network. Based on their queuing packets, node n_i has a set of candidate streams S_i, where the destination node of the q-th stream $s_{iq} \in S_i$ is denoted as $d(s_{iq})$. Let the parameter set $\{y_{iq}\}$ ($y_{iq} \in \{0,1\}, i = 1, \ldots, N_n, q = 1, \ldots, |S_i|$) denote whether the q-th candidate stream of node i is transmitted in the current TD. If a stream s_{iq} is transmitted, $y_{iq} = 1$; otherwise, $y_{iq} = 0$. Similarly, $\{t_i\}$ and $\{h_i\}$ ($t_i, h_i \in \{0,1\}, i = 1, \ldots, N_n$) are used to denote the transmitter and receiver node assignment in the current TD respectively. If node n_i is selected as a transmitter/receiver node, we have $t_i = 1/h_i = 1$, otherwise $t_i = 0/h_i = 0$. If $t_i = h_i = 0$, node n_i is recognized as an idle node. The assignment of a stream to a specific antenna of a transmitter is represented by the parameter a_{iqk} ($a_{iqk} \in \{0, 1\}, i = 1, \ldots, N_n, q = 1, \ldots, |S_i|$ and $k = 1, \ldots, N_i^{ant}$), where $a_{iqk} = 1$ if stream s_{iq} is assigned to transmit from antenna k of node n_i. The transmission rate of stream s_{iq} is impacted by both the strength of the stream $(i, d(s_{iq}), k)$ (denoted as $\mathcal{S}(s_{iq})$), and the interference level at receiver node $d(s_{iq})$ (denoted as $\mathcal{I}(d(s_{iq}))$). The priority of stream s_{iq} depends on the priority of its associated packet and is denoted as $\mathscr{P}(s_{iq})$.

The scheduling process selects a set of streams to transmit among all the candidate ones in the current TD. The objective of the scheduling is to maximize the sum of priority-weighted capacity of the scheduled streams, so that both data rate and priority can be jointly optimized. The problem is formulated as follows:

$$\max U = \sum_{n_i \in N} \sum_{s_{iq} \in S_i} y_{iq} \mathscr{C}(\mathcal{S}(s_{iq}), \mathcal{I}(d(s_{iq}))) \mathscr{P}(s_{iq}); \quad (4)$$

$$\sum_{s_{iq} \in S_i} a_{iqk} \leq 1, i = 1, 2, \ldots, N_n, k = 1, \ldots, N_i^{ant}; \quad (5)$$

$$\sum_{s_{iq} \in S_i} y_{iq} \leq \mathcal{N}_i^{tx}, i = 1, 2, \ldots, N_n; \quad (6)$$

$$h_i \sum_{m \in \mathscr{V}_i} \sum_{\substack{s_{mq} \in S_m \\ d(s_{mq}) = i}} y_{mq} + h_i \sum_{m \in \mathscr{V}_i} \sum_{\substack{s_{mq} \in S_m \\ d(s_{mq}) \neq i}} y_{mq} \leq \mathcal{N}_i^{rc},$$
$$i = 1, 2, \ldots, N_n; \quad (7)$$

$$t_i + h_i \leq 1, i = 1, 2, \ldots, N_n; \quad (8)$$

$$a_{iqk} \leq y_{iq} \leq t_i, a_{iqk} \leq y_{iq} \leq h_{d(s_{iq})},$$
$$i = 1, \ldots, N_n, q = 1, \ldots, |S_i|, k = 1, \ldots, N_i^{ant}; \quad (9)$$
$$t_i, h_i, y_{iq}, a_{iqk} \in \{0, 1\}.$$

Constraint (5) ensures that an antenna can only transmit one scheduled stream at most in each slot. Equation (6) constrains the total number of transmitted streams from n_i should be no more than its transmitting constraint value $\mathcal{N}_i^{tx}$, which depends on the antenna numbers of n_i and all its neighboring receivers as well as the channel independency level between n_i and every receiver. Equation (7) provides the constraint at receiver n_i where the total number of receiving streams including data streams (the first term on the left side) and interference streams (the second term on the left side) is restricted to be no more than its receiving constraint value $\mathcal{N}_i^{rc}$ in order to decode the receiving packet. Equation (8) represents that nodes in the network are half-duplex; and equation (9) ensures the parameters to have the correct relationship. So far, we formulate the problem of heterogeneous stream scheduling as an integer programming problem with the objective function in (4) subject to constraints (5)-(9).

Note that the the strength of a stream $\mathcal{S}(s_{iq})$ will reduce if more streams are scheduled to transmit from n_i, which will be incorporated during the stream scheduling process. As the interference $\mathcal{I}(d(s_{iq}))$ will not be known until the scheduling is completed, we will use an average interference level estimated from the past transmissions. In addition, a receiver cannot cancel the interference when the total number of streams it receives is beyond its decoding capability or it does not have channel knowledge, or the interference is due to decoding errors as a result of inaccurate channel knowledge or a-synchronization [9]. The last two types of interference is included in the measured interference. As our MAC design ensures that the number of concurrent transmissions from one-hop transmitters is below the decoding capability of each receiver and with channel estimation, so the un-cancelable interference is from transmitters two hops or more away and is thus weaker. Also, the number of antennas of nodes in the ad-hoc network is generally small, so the number of steps needed for interference cancelation and the error propagation is also limited. Based on the actual decoding quality, the estimated interference level can be adjusted, and set higher to select stronger streams for more reliable decoding at the cost of possible reduction of the number of concurrent streams. Further, as our algorithms schedule stronger streams, it helps to significantly increase the signal to interference plus noise ratio and mitigate the interference impact.

Proposition I: The heterogeneous stream scheduling (HSS) problem described above is NP-hard.

Proof: First we introduce a simplified version of HSS problem represented by a graph $G = (V, E)$. A vertex $v_i \in V$ represents a node n_i, and an edge $e = (v_i, v_k)$ denotes that n_i and n_k are neighbors in the network. Assume each node has a candidate stream s for each of its neighbors, and the gain of scheduling $\mathscr{C}(s)\mathscr{P}(s)$ is 1 for all s. The transmitting and receiving constraints for all n_i are $\mathcal{N}_i^{tx} = \mathcal{N}_i^{rc} = 1$. The optimum scheduling solution of the simplified HSS problem is a maximum set of vertices that can transmit simultaneously while $\mathcal{N}_i^{tx}$ and $\mathcal{N}_i^{rc}$ are satisfied for transmitter and receiver nodes respectively. The simplified HSS problem can be proved to be NP-hard by reducing the NP-complete maximum independent set (MIS) problem to it. For any instance of MIS represented by a graph $G' = (V', E')$, form a new graph $G = (V, E)$ in the following way. Keep the vertex set V' and replace each edge in E' with a dummy vertex, denoted as a set V_d, so that $V = V' \bigcup V_d$. Connect each dummy vertex in V_d to the two original end vertices in V'. The dummy vertices that represent edges connected to the same vertex in G' are also connected in G. It is then straightforward to see that the optimum scheduling solution of the simplified HSS problem in G gives an equivalent solution of MIS problem in G'.□

5 Centralized Algorithm

Due to the NP-hardness of the problem, an efficient heuristic algorithm is required to solve the scheduling problem. In algorithm 1, we propose a centralized algorithm. In lines 1-7, a set W is constructed to include all the candidate streams from every node in the network. In lines 8-17, the centralized algorithm greedily schedules the stream with the highest weight for transmission in a TD, while meeting the constraints in equations (5)-(9). In line 11-12, the selected stream is assigned to be transmitted from the corresponding transmitter to the receiver. As a node cannot be a transmitter or receiver at the same time, in line 13, all the candidate streams that have transmission conflict with the scheduled stream $s = (i^*, d(s_{i^*q^*}), k^*)$ are removed from the set W, including the candidate streams that have the node n_{i^*} as the receiver, have $n_{d(s_{i^*q^*})}$ as the transmitter, or have node n_{i^*} as the transmitter and are associated with the antenna k^*. It can be proved that the centralized algorithm is within a fixed approximation ratio to the optimal solution [22].

Algorithm 1 Centralized Scheduling

Initialize: $W \Leftarrow \oslash$
for $i = 1$ to N_n **do**
  if $\exists s_{iq}, \forall q \in \{1, \ldots, |S_i|\}$ **then**
    $w(iqk) \Leftarrow \mathscr{C}(iqk)\mathscr{P}(iq), \forall k \in \{1, \ldots, N_i^{ant}\}$
    $W \Leftarrow W \bigcup \{w(iqk)\}$
  end if
end for
while $W \neq \oslash$ **do**
  $(i^*, q^*, k^*) = \arg\max_{\{i,q,k\}} W$, the corresponding destination node is $d(s_{i^*q^*})$
  if Selecting $(i^*, d(s_{i^*q^*}), k^*)$ satisfies constraints (6) and (7) **then**
    Assign $n_{i^*}/n_{d(s_{i^*q^*})}$ as the transmitter/receiver node
    Schedule the stream $(i^*, d(s_{i^*q^*}), k^*)$
    $W \Leftarrow W \setminus \{w(iqk) | \forall i, q \ \ s.t. \ \ d(s_{iq}) = i^*, k = 1, \ldots, N_i^{ant}\} \bigcup \{w(iqk) | i = d(s_{i^*q^*}), q = 1, \ldots, |S_i|, k = 1, \ldots, N_i^{ant}\} \bigcup \{w(iqk) | i = i^*, k = k^*, q = 1, \ldots, |S_{i^*}|\}$
  else
    $W \Leftarrow W \setminus w(i^*q^*k^*)$
  end if
end while
for nodes in T **do**
  if streams are towards the same receiver **then**
    precoding
  end if
end for

Proposition 2: The centralized scheduling algorithm can achieve an approximation ratio of $1/\left((2 + \mathcal{D}) \max_i\{N_i^{ant}\} + 2\right)$, where $\mathcal{D}$ is the maximum node degree in the network.

The centralized algorithm with the proved approximation ratio serves as a benchmark for performance comparison. From the formulation (4)-(9), it is clear that the scheduling problem has to determine the values of the parameter sets: $\{t_i\}$, $\{h_i\}$, $\{y_{iq}\}$ and $\{a_{iqk}\}$ to assign a packet to an appropriate transmitter antenna in order to maximize the total weighted rate of the network. In a practical distributed half-duplex network, it is reasonable to divide the problem into two parts: transmitter selection and stream allocation, where the first phase determines the values of $\{t_i\}$ and $\{h_i\}$, and the second phase determines the value of $\{y_{iq}\}$ and $\{a_{iqk}\}$ to assign a packet to a specific transmission stream. In the next section, the two subproblems are solved separately.

6 Distributed Algorithm and Protocol

In order to address the network heterogeneity, our algorithm groups transmissions into two types, transmissions to poor nodes using *P-slots* and to rich nodes using *R-slots*. The current slot type is determined in a distributed manner by each node and the nodes in a neighborhood reach a consensus through signaling exchange. In both types of slots, spatial multiplexing, selection diversity and transmitter precoding are adaptively utilized to deal with varying traffic demands and channel conditions to improve the overall network performance.

The distributed scheduling algorithm consists of two phases, namely *transmitter node selection / slot request* and *stream allocation*. In the first phase, a set of nodes are first selected to be transmitter nodes, and each node differentiates its packets for poor nodes and rich nodes to determine its current preference of transmission slot type. In the second phase, stream allocation is performed to allocate the data packets of the transmitter nodes to a selected set of antennas with an appropriate MIMO strategy.

In the rest of this section, we first present our scheduling algorithm in sequence of the two phases mentioned above. The complete protocol is then introduced, where we explain the detailed procedures taken to implement the algorithm and calculate the required parameters in a distributed environment.

6.1 Transmitter Node Selection and Slot Request

In this phase, nodes are distributively selected as transmitter nodes and their preference of slot type is decided. Instead of randomly selecting the transmitter nodes, the transmitter selection phase supports service differentiation and reduces transmission delay by giving a higher transmission priority to the streams that are with packets in higher service class and/or have larger queuing delay. Additionally, the type of transmission slots is differentiated to support transmissions to heterogeneous nodes. We first give the main idea and define parameters used for the selection, then we discuss the details of the selection process.

6.1.1 Basic Plot

In MIMO transmissions, in order to not exceed the decoding capacity of nodes, the number of streams that can be simultaneously transmitted in a neighborhood is constrained. Therefore, the number of transmitter nodes selected in our algorithm also has a limit, which will avoid unnecessary channel measurement. In addition, the decoding capabilities of receivers, represented by their receiving constraints in Section 4.1, are different in a heterogeneous MIMO network. In our algorithm, each node distributively determines if it can serve as a transmitter node in a transmission duration, and selects the type of slot used for transmission based on the decoding capacity of its neighboring receivers.

Based on the receiving constraint, an active node n_i which has data to send groups its neighboring nodes into poor node set $\mathscr{V}_i^p$ and rich node set $\mathscr{V}_i^r$ based on the receiving constraint $\mathcal{N}_k^{rc}$ of a neighbor n_k, which is broadcast with the Hello messages sent periodically at the network layer. We introduce a threshold value $\mathcal{T}_i^{TX}$, which is calculated separately for each of the two sets. Denote the set of neighboring nodes *in concern* as $\mathscr{V}_i$, where $\mathscr{V}_i$ can correspond to $\mathscr{V}_i^p$ or $\mathscr{V}_i^r$ depending on which set is concerned at the calculation time. The parameter $\mathcal{T}_i^{TX}$ of n_i is estimated based on the number of active nodes around a neighboring node $n_j \in \mathscr{V}_i$ (denoted as N_j^{active}) and the receiving constraint of node n_j (denoted as $\mathcal{N}_j^{rc}$) as $\mathcal{T}_i^{TX} = \min\{1, \min_{j \in \mathscr{V}_i}\left(\mathcal{N}_j^{rc}/N_j^{active}\right)\}$. To support some transmission fairness, the neighboring transmitters of n_j can be evenly allocated the transmission opportunities based on the decoding constraint of n_j. Therefore, $\mathcal{T}_i^{TX}$ represents the probability of a node n_i being a transmitter in order to ensure all neighbors in $\mathscr{V}_i$ to perform the correct decoding. A node n_i can be selected as a transmitter if the value of an appropriately calculated random variable is below $\mathcal{T}_i^{TX}$.

Recall that we use stream priority to represent how urgent a stream transmission is. It is therefore natural to use the average stream priority to reflect the level of priority for a node to be a transmitter. Denote all candidate streams (i.e. the head-of-queue packets with the number constrained by the number of antennas of n_i) of n_i as a set S_i and the priority of a stream s_{iq} as $\mathscr{P}(s_{iq})$, the priority of a node n_i can be represented by the average priority of its candidate streams as $\mathscr{P}_i = \sum_{s_{iq} \in S_i} \mathscr{P}(s_{iq})/|S_i|$. A node n_i can calculate the average priority $\bar{\mathscr{P}}_i$ of all the N_i^{active} active nodes in its neighborhood as $\bar{\mathscr{P}}_i = \sum_{j=1}^{N_i^{active}} \mathscr{P}_j / N_i^{active}$. 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 nodes not having packets sent in a TD can be predicted as time moves forward.

To avoid extra signaling and control overhead, an active node n_i *self-decides* if it should be selected as a transmitter node by calculating an index number $\mathcal{X}_i^{TX} = (\bar{\mathscr{P}}_i - \hat{\mathscr{P}}_i)/\bar{\mathscr{P}}_i + \gamma_i$. Here the parameter γ_i is a random number uniformly distributed in the range [0,1] and generated by a node n_i at each transmission duration (TD) to provide some fairness among nodes. $\hat{\mathscr{P}}_i$ is the average priority of candidate streams at node n_i that are targeted for nodes in $\mathscr{V}_i$. The factor $(\bar{\mathscr{P}}_i - \hat{\mathscr{P}}_i)/\bar{\mathscr{P}}_i$ is used to give the higher priority node a larger probability for transmission. In a TD, if $\mathcal{X}_i^{TX} < \mathcal{T}_i^{TX}$, node n_i is selected as a transmitter node for receiver nodes in $\mathscr{V}_i$; otherwise, it has no right of transmission. Our transmitter selection algorithm prefers a node with a higher service level and/or a larger load and hence longer delay, and thus supports QoS and load balancing while ensuring certain fairness. Our selection is conservative as it considers the decoding capability of all the neighboring nodes instead of only that of the actually selected receiver nodes known only after the scheduling.

6.1.2 Selection Process

To give priority to transmissions towards poor nodes, at the beginning of a transmission duration, an active node n_i first determines whether it needs to initiate a transmission using P-slot based on the priority of its streams targeted for poor nodes in $\mathscr{V}_i^p$. For the subset of candidate streams in S_i destined to poor nodes in $\mathscr{V}_i^p$, their average priority can be calculated as $\mathscr{P}_i^p = (\sum_{k \in \mathscr{V}_i^p} \sum_{m \in S_{i,k}} \mathscr{P}(m)) / \sum_{k \in \mathscr{V}_i^p} |S_{i,k}|$, where $S_{i,k}$ is the set of candidate streams from node n_i to the poor node n_k. Let $\mathscr{V}_i = \mathscr{V}_i^p$ and substitute $\hat{\mathscr{P}}_i$ by $\mathscr{P}_i^p$ for calculating the index $\mathcal{X}_i^{TX}$, which is compared with $\mathcal{T}_i^{TX}$ calculated based on nodes in $\mathscr{V}_i^p$. If $\mathcal{X}_i^{TX} < \mathcal{T}_i^{TX}$, node n_i can be a transmitter node and initiate a P-slot transmission. The P-slot streams are selected so that the receiving constraints are satisfied at a targeted poor receiver. Otherwise, node n_i checks if it can be a transmitter using R-slot. Similar to the previous step, $\mathcal{T}_i^{TX}$ is calculated concerning nodes in $\mathscr{V}_i^r$ and $\mathcal{X}_i^{TX}$ is obtained by letting $\hat{\mathscr{P}}_i$ equal to $\mathscr{P}_i^r = (\sum_{k \in \mathscr{V}_i^r} \sum_{m \in S_{i,k}} \mathscr{P}(m)) / \sum_{k \in \mathscr{V}_i^r} |S_{i,k}|$, where $S_{i,k}$ is the set of candidate streams which are from node n_i to a rich node n_k. Node n_i is selected as a transmitter node for receiver nodes in $\mathscr{V}_i^r$ if the updated parameters satisfy $\mathcal{X}_i^{TX} < \mathcal{T}_i^{TX}$.

If a node determines to be a transmitter node, it broadcasts an RTS message indicating the slot type as discussed in 6.3. After the transmitters and the slot types are confirmed by the receiver nodes through CTS transmission, the transmitter nodes proceed to the second phase of the scheduling described next.

6.2 Stream Allocation

Stream allocation is performed distributively at each of the selected transmitter nodes. The selection gives preference to streams with higher priority. For streams of the same priority, to achieve a higher data rate, the allocation process is solely based on the stream capacity by *opportunistically* assigning a channel with good condition to a selected stream. For a high-priority stream that does not have high-quality channel, the selection process reserves more of the total transmitting power for the stream to ensure a higher transmission reliability.

For a selected transmitter, there is a limit on the number of streams it is allowed to transmit, in order to meet the receiving constraints at all neighboring receivers. For a selected transmitter n_i, let N_i^0 be the number of pre-selected streams to be transmitted and N_i^{allo} be the number of streams node n_i is allowed to transmit, which is calculated based on feedbacks from neighboring receivers as described in Section 6.3. Suppose the N_i^0 candidate streams have L_i distinct priority levels. The receiver nodes that the candidate streams are targeted for are then partitioned into subsets $\{D_i^1\}, \{D_i^2\}, \ldots, \{D_i^{L_i}\}$ according to the descending priorities of the streams, where the set $\{D_i^j\}$ contains the target receiver nodes of the streams with the j-th highest priority level, and the q-th element in $\{D_i^j\}$ is denoted as $D_i^j(q)$. Recall that a stream s is identified by its transmitter node, receiver node and transmitter antenna. Denote the set of antennas that node n_i has as $\{A_i\}$, and the p-th element is $A_i(p)$. For a stream of n_i which has the receiver $D_i^j(q)$ and transmitting antenna $A_i(p)$, the stream capacity $\mathscr{C}(i, D_i^j(q), A_i(p))$ depends on the stream strength and the estimated interference level at the receiver node $D_i^j(q)$, as discussed in Section 4.3. For transmitter node n_i, there is a set W_i^0 consisting of all the capacity parameters of the candidate streams $W_i^0 = \bigcup_{j=1}^{L_i} \{\mathscr{C}(i, D_i^j(q), A_i(p)) | A_i(p) \in \{A_i\}, D_i^j(q) \in \{D_i^j\}, p = 1, \ldots, |\{A_i\}|, q = 1, \ldots, |\{D_i^j\}|\}$.

The procedure of stream allocation is described in the algorithm 2, where j is the index of the priority level, $\{A_i\}^{res}$ is the set of remaining available antennas of node n_i and N_i^{res} is the residual number of streams to allocate. The initial value of N_i^{res} is set to be the total number of streams for allocation N_i^{allo}. As in lines 2-11, the algorithm starts from the set of candidate streams which have the highest priority ($j = 1$), and calls the subroutine *OPPORTUNISTIC_ALLOCATION* as in algorithm 3 for each priority level, until all the allowed streams have been allocated or the antennas of node n_i have all been assigned or reserved for streams. In lines 12-16, power is allocated to the selected antennas based on the transmission pattern. As described in section 4.1, precoding is used to maximize the data rate between a node pair if all streams are scheduled to transmit towards the same receiver where optimal power allocation is performed through water-filling; when streams are towards different receivers, power is simply distributed evenly among the antennas.

The subroutine *OPPORTUNISTIC_ALLOCATION* is described in algorithm 2 to allocate k antennas to transmit the streams of the j-th highest priority level that are targeted for the receiver set $\{D_i^j\}$. The parameter N_i^{res} is the residual number of antennas available for allocation, the set $\{A_i\}^{res}$ contains the candidate antennas of node n_i for stream allocation, W_i^j contains the capacity parameters of the streams formulated between the antennas in $\{A_i\}^{res}$

Algorithm 2 Distributed Scheduling

Initialize: $j = 1$, $\{A_i\}^{res} = \{A_i\}$, $N_i^{res} = N_i^{allo}$
while $N_i^{res} > 0$ **do**
  if $|\{D_i^j\}| \leq N_i^{res}$ **then**
    OPPORTUNISTIC_ALLOCATION$(\{A_i\}^{res}, \{D_i^j\}, |\{D_i^j\}|, N_i^{res})$
    $N_i^{res} = N_i^{res} - |\{D_i^j\}|$
  else
    OPPORTUNISTIC_ALLOCATION$(\{A_i\}^{res}, \{D_i^j\}, N_i^{res}, 0)$
    $N_i^{res} = 0$
  end if
  $j \Leftarrow j + 1$
end while
if All streams are towards one receiver **then**
  Use precoding and optimal power allocation
else
  Power is evenly distributed
end if

Algorithm 3 *OPPORTUNISTIC_ALLOCATION* $(\{A_i\}^{res}, \{D_i^j\}, k, N_i^{res})$

1: **Initialize:** $l = 0$, $W_i^j = \{\mathscr{C}(i, D_i^j(q), A_i^{res}(p)) | A_i^{res}(p) \in \{A_i\}^{res}, D_i^j(q) \in \{D_i^j\}, p = 1, \ldots, |\{A_i\}^{res}|, q = 1, \ldots, |\{D_i^j\}|\}$
2: **while** $l < k$ **do**
3:   $W_{max} \Leftarrow \max W_i^j$, $\{A_{max}, D_{max}\} \Leftarrow \arg\max W_i^j$
4:   Allocate the stream for the receiver D_{max} to the antenna A_{max};
5:   $W_i^j \Leftarrow W_i^j \setminus \{W(A_{max}, D_i^j(q)) | D_i^j(q) \in \{D_i^j\}, q = 1, \ldots, |\{D_i^j\}|\}$; if there is no other stream target for the receiver node D_{max}, also remove $\{W(A_i^{res}(p), D_{max}) | A_i^{res}(p) \in \{A_i\}^{res}, p = 1, \ldots, |\{A_i\}^{res}|\}$;
6:   **if** D_{max} has sent indicator of weak channel **then**
7:     **if** $N_i^{res} > 0$ **then**
8:       $k \Leftarrow k - 1$, $l \Leftarrow l + 1$, $N_i^{res} \Leftarrow N_i^{res} - 1$;
9:     **else** $\{N_i^{res} = 0\}$
10:       $k \Leftarrow k - 1$
11:     **end if**
12:   **end if**
13:   $\{A_i\}^{res} \Leftarrow \{A_i\}^{res} \setminus A_{max}$
14:   $l \Leftarrow l + 1$
15: **end while**

and the receivers in $\{D_i^j\}$ and l represents the number of streams currently allocated. The allocation is based on spatial multiplexing and selection diversity, and in sequence of descending stream quality. As the allocation scheme favors stream priority than stream quality, in some cases, although the channel condition is severe, a transmission with a high priority is still permitted. To reduce erroneous decoding thus packet loss under the severe channel condition, when a selected stream does not have good enough quality as indicated by a weak channel indicator include in the CTS (Section 6.3), the total number of antennas available for allocation of this stream is decreased by one to reserve extra transmitting power for the weak stream to improve its quality, as in lines 6-12.

6.3 Implementation of Distributed Scheduling

To enable the proposed many-to-many transmission and better exploit various diversity techniques for higher capacity and reliability, the implementation of the distributed scheduling algorithm is TDMA-based, where the time is divided into a serials of transmission duration consisting of four phases with different lengths. The duration of each phase is fixed and enough for the corresponding message transmission. Following the convention of IEEE 802.11 DCF, signaling messages are named RTS, CTS, DATA and ACK, which are transmitted during phase I, II, III and IV respectively. Note that slot synchronization is currently achievable in the IEEE802.11 family of protocols. By taking advantage of the selection diversity and multi-user diversity, our scheme could effectively increase the SINR of a received signal, which would help improve the accuracy of synchronization as well as mitigate the impact of a-synchronicity in a distributed scenario. The procedure of signal exchange and information acquisition for heterogeneous MIMO scheduling is as follows.

Phase I: Transmission Request and Slot Conservation. At the beginning of phase I, a node n_i which selects itself as a transmitter node as in Section 6.1 broadcasts an RTS. Before sending out the RTS, node n_i selects a set of highest-priority data packets from its queue to form $N_i^0 \leq N_i^{max}$ candidate streams, where N_i^{max} is the maximum number of streams that can be transmitted in a transmission duration depending on the number of antennas of n_i, and the amount of data queued. The IDs of the target receiver nodes of the selected packets, the value N_i^0, as well as the ID of node n_i are then included in the RTS. If n_i wants to request a P-slot towards node n_k, an RTS should further carry an indicator of P-slot and the calculated average priority $\mathscr{P}_i^p$.

The preamble of a packet is used as the training sequence for the channel estimation purpose. After the RTS is transmitted from the first antenna of the transmitter node, for both types of slot, the preamble is rotationally broadcasted through the remaining antennas of the transmitter node with a short notice signal separating two antennas' transmissions, so that the spatial channels between each antenna of the transmitter nodes and the receiver nodes can be differentiated and estimated. An RTS is masked by another random code, called ID code, which are almost orthogonal for different nodes and assigned similarly to that in [23], so a receiver node can get the channel information of different transmitter nodes from concurrently received RTSs. Our transmitter node selection algorithm in Section 6.1 adaptively selects a subset of nodes in a neighborhood to participate in channel estimations based on the decoding capabilities of nodes in the neighborhood, which not only reduces the channel estimation complexity and avoids unnecessary channel estimations but also constrains the total interference in a neighborhood for better decoding.

Phase II: Transmission Confirmation. Upon receiving multiple RTSs, a receiver correlates its received signals with each element in its set of random codes to differentiate the training sequences from different transmitter nodes, estimates spatial channels and extracts other information included in RTSs.

If a node n_k receives a request for P-slot transmission to itself, it sorts all P-slot requests it receives (for itself or for other receiver nodes) based on the request priorities. When multiple requests have the same priority, the request for the receiver with a higher ID is preferred. The receiver n_k then checks the number of P-slot transmissions allowed in the neighborhood from higher priority to lower priority until all the requests are accommodated or n_k is fully-loaded with data and/or interference streams. Denote the number of P-slot requests accommodated at node n_k as $N_{k,p}^{dec}$, which does not exceed the receiving constraint of n_k, $\mathcal{N}_k^{rc}$. If n_k is a target receiver of some of the accommodated requests, it considers the current transmission duration as P-slot and broadcasts a CTS with its list of confirmed P-slot requests.

If n_k is only the target receiver of some R-slot requests, while it may overhear some P-slot requests for other receivers, it checks whether it has enough residual stream $N_{k,r}^{dec} = \max\{0, \mathcal{N}_k^{rc} - N_{k,p}^{dec}\}$ for R-slot transmission. If $N_{k,r}^{dec} > 0$, it considers the current transmission duration as R-slot. Different from P-slot transmission in which a transmitter node pre-selects a target receiver, transmission streams are flexibly selected for different receivers in R-slot based on the channel condition to improve the aggregate data rate. After node n_k decodes the information in RTSs from all the selected transmitter nodes in its neighborhood, it learns the number of R-slot streams it may receive in the current duration, $N_{k,r}^0$, including the data streams targeted to itself and the interference streams targeted to other nodes. Denote all transmitter nodes in the one-hop neighborhood of n_k as $\mathscr{V}_k^t$, and each transmitter n_j requires N_j^0 R-slot streams for transmission, we have $N_{k,r}^0 = \sum_{j \in \mathscr{V}_k^t} N_j^0$. Node n_k then broadcasts $N_{k,r}^0$ and $N_{k,r}^{dec}$ through CTS.

A stream may have poor quality, when there is a long distance or deep-fading channel between a transmitter and a receiver. A receiver estimates the strength of a data stream based on the signal-to-noise-ratio (SNR) of the training signal. If the received SNR is lower than a threshold, it includes a weak-channel indicator in the CTS to inform the transmitter to select a more reliable transmission scheme.

To allow the transmitter to estimate the spatial channels to the receiver, the preamble of CTS is utilized as a short training sequence and rotationally broadcast from node n_k's antennas $1 \sim N_k^{ant}$, as in the case of RTS. A CTS signal is also masked by the ID code of n_k.

Phase III: Stream Allocation and Transmission. By differentiating multiple CTSs and extracting the information included, a node n_i estimates the channel matrix $\mathbf{H}_{ki}$ between itself and each active receiver node n_k, and obtains its transmitting constraint value $\mathcal{N}_i^{tx}$. Specifically, if node n_i sends out a P-slot request in the RTS phase, it checks if its P-slot request has been confirmed by all the CTSs. Denote the number of streams confirmed by CTSs as $\tilde{N}_i^0$, so the number of streams allowed for transmission can be calculated as $N_i^{allo} = \min\{\mathcal{N}_i^{tx}, \tilde{N}_i^0\}$. Node n_i allocates the stream following the procedure in 6.2 according to the estimated spatial channels.

If n_i receives a confirmation for its R-slot request, it has to determine N_i^{allo} based on the total R-slot confirmations included in the CTSs from rich neighboring receivers in the set $\mathscr{V}_i^r$. Each responding receiver n_k sends back the total number of streams it may receive, $N_{k,r}^0$, the maximum number of streams it can decode, $N_{k,r}^{dec}$, and possibly weak-channel indicators. In order to ensure all the receiver nodes in its neighborhood to have a high probability of correct decoding, node n_i constrains its number of sending streams to a rounded integer number as $N_i^{allo} = \min\{\mathcal{N}_i^{tx}, N_i^0 \min_{k=1}^{N_i^r} \left(N_{k,r}^{dec}/N_{k,r}^0\right)\}$.

With the estimation of all spatial channels between n_i and its target nodes, the set W_i^0 of stream capacity factors is constructed and the stream allocation described in algorithm 2 is then performed to transmit the data streams through the selected antennas. Meanwhile, receiver nodes decode streams from the neighboring transmitter nodes using channel coefficients estimated in phase I.

Phase IV: Acknowledgement. If a data stream is decoded correctly, the receiver node has to confirm the reception. An ACK is masked with the ID code of the receiver and broadcast, carrying the IDs of the transmitter nodes whose streams have been correctly received.

In phase IV, all transmitter nodes are in listening mode. A transmitter node extracts the information in ACKs and removes the correctly received data packets from the queue, and keeps the erroneously received or lost data packets in the queue for scheduling in the next transmission duration.

Note that random ID codes are only used for differentiation in control signal transmission. As control signals are relatively short and sent at the maximum power, there is no significant overhead induced for packet encoding and decoding and there is no need for power control.

6.4 Examples

In this section, we give two examples to illustrate the process of the stream allocation algorithm based on the simple topology as in Fig. 1.

Suppose that each of nodes 2 and 3 has packets for both nodes 1 and 4. Depending on the antenna array sizes, both 2 and 3 regard node 1 as a poor node and node 4 as a rich node. Following the transmitter selection scheme, both node 2 and 3 may select themselves to be transmitters. As poor nodes have higher priority to receive packets, node 2 and 3 may both initiate P-slot transmission towards node 1 in the first TD. As node 1 can only receive one stream, it confirm the request from the node with the higher priority, say it is node 3. To avoid interference, node 2 cannot transmit at this TD. In the second TD, node 2 still initiates a P-slot transmission to node 1. After confirmed by node 1, one stream is transmitted in this TD. In the third TD, both 2 and 3 start to initiate R-slot transmission towards node 4, each with two streams in the RTS request. Node 4 then feeds back $N_{4,r}^0 = 4$ and $N_{4,r}^{dec} = 3$ in the CTS. Following the procedures of Phase III, nodes 2 and 3 transmit three streams in total, i.e. two from node 2 and one from node 3, all towards node 4. As the streams from each sender are transmitted towards one receiver in each TD, the streams are precoded with the power allocated optimally among transmitting antennas. Note that in the counterpart algorithms where heterogeneity is not considered, nodes 2 and 3 can only transmit one stream in total in each TD, as node 1 is an active receiver and it always restricts the number of streams in the neighborhood.

Consider another scenario for the same topology that only nodes 1 and 4 have packets for nodes 2 and 3. As node 2 and 3 are considered as poor nodes by node 1 and rich node by node 4, node 1 first initiates P-slot transmission towards 2 and 3 respectively in two consecutive TDs. As node 1 only has one antenna, it can only transmits one stream in each TD. As a result, node 2 and 3 still have one DoF which can be used to receive a stream from node 4. This stream is selected by node 4 as described in 6.2, and the stream with the highest capacity in the first candidate stream subset $\{D_4^1\}$ is selected. So there are 2 streams transmitted in the network in each of the first two TDs. Suppose that node 4 still has packets for 2 and 3 in the following TD, R-slot transmission requests are thus sent towards node 2 and node 3 simultaneously. If the channels are rich-scattered, the DoF of the channel between 4 and any of the two receivers is $\min\{N_4^{ant}, N_2^{ant}\} = \min\{N_4^{ant}, N_3^{ant}\} = 2$. According to the transmitting degree constraint, only two streams can be transmitted. Node 4 therefore selects two candidate streams following the procedure in 6.2. If the two streams selected are towards two different receivers 2 and 3 respectively, power is distributed evenly over the two antennas; if the two streams selected are towards the same receiver, precoding and optimal power allocation are assumed. If there is strong LOS component presented between node 4 and node 1, and $DoF(4,1) = 1$, we have $\mathcal{N}_4^{tx} = 1$ and node 4 can only transmit one stream as a result. The stream selection is done by node 4 similarly as that in the first two TDs.

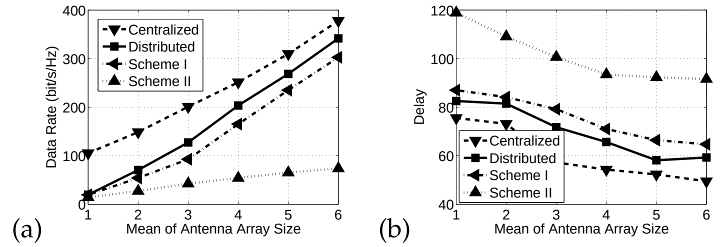

Fig. 3. Impact of mean of antenna array size: (a) data rate; (b) delay.

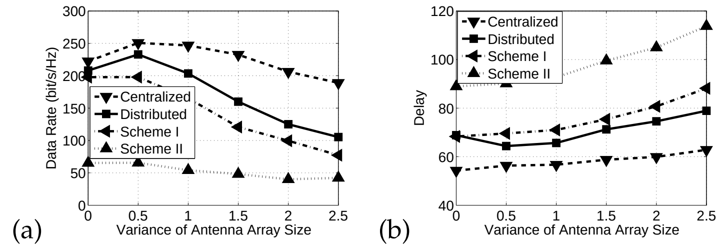

Fig. 4. Impact of variance of antenna array size: (a) data rate; (b) delay.

7 Performance Evaluation

In this section, we evaluate the performance of our proposed algorithms through simulations. Nodes are distributed uniformly over a $1250m \times 1250m$ area and form an ad-hoc network with random topology. Each node has a transmission range of $250m$. The bandwidth of the channel is $20MHz$, and the length for data slot of a TD is $2.5ms$ so the MIMO channel can be considered as quasi-static during a TD [24]. To model a heterogeneous MIMO ad hoc network, we assume the antenna array sizes of nodes in the network are normally distributed with a given mean and variance. The channel is modeled based on the antenna array sizes, the distance between nodes and the small-scale fading coefficients following Rayleigh/Ricean model. The incoming traffic is Poisson distributed with a given mean value λ and the sources and destinations are chosen at random. The size of a packet is 1000 bytes. White Gaussian noise with $SNR = 10dB$ is added to include environment noise and interference that cannot be canceled. A result is obtained by averaging over 10 runs of simulations with different random seeds.

The distributed scheduling algorithm proposed in Section 6, including both transmitter nodes selection 6.1 and stream allocation 6.2, is implemented based on the protocol described in Section 6.3. Compared with conventional scheduling strategies in MIMO ad hoc networks, our distributed algorithm has the following unique features: adaptive use of different transmission strategies based on node types and channel conditions, and enabling multi-user to multi-user transmissions exploiting both cooperative multiplexing and selective diversity. To demonstrate the benefits of these features, we design two alternative schemes here for reference. *Scheme I* is based on the opportunistic and cooperative spatial multiplexing scheme proposed in [10], which supports many-to-many cooperative transmission, but does not have specific strategies to handle the heterogeneity of nodes and channels. *Scheme II* takes the conventional scheduling strategy in MIMO ad hoc networks, where during each TD only one pair of transmitter/receiver nodes is allowed to communicate in a neighborhood with as many streams as possible. In each transmission duration, the node pair with the best channel quality is selected, and transmitter node selection is also implemented here to reduce collision. To provide a benchmark for performance comparison, we also implemented the centralized scheduling algorithm proposed in Section 5.

The metrics we use for comparison are aggregate data rate and average delay. Aggregate data rate is the total data rate of the network averaged over the number of transmission durations. Average delay is the average number of transmission durations a packet waits in the queue before it is successfully transmitted. It is defined in Section 4.1 that the stream priority depends on the service type and queuing delay. However, as our focus is to study the effectiveness of the scheduling algorithms with delay as one of the metrics, the incoming packets are assigned with the same service priority and the priority of a packet increases as the delay becomes larger. We investigate the impact of a set of factors on performance, namely the mean value and variance of antenna array size, the LOS component, node density, traffic arrival rate and mobility. For each factor, the centralized algorithm and distributed algorithm as well as the two reference schemes are implemented, and both data rate and average delay are compared. If not otherwise specified, the value of K-factor is 0, the number of nodes in the network is 100, the mean and variance of degree-of-freedom are 4 and 1 respectively, the average packet arrival rate λ is 5 packets per link and the network is static.

7.1 Impact of the Mean of Antenna Array Size

The mean value of antenna array size determines the average node transmission capability, and it impacts the overall capacity of the network. In Fig. 3, as the mean value grows, all schemes except scheme II obtain significantly higher data rate and lower delay, as these schemes can better exploit the multiplexing gain and diversity gain allowed by a larger antenna array to form many-to-many transmissions with a larger number of selected streams. Particularly, the formulation of cooperative multiplexing transmissions alleviates the limitation of the number of transmit antennas. In contrast, the performance of scheme II is constrained by both the number of transmit antennas and receive antennas, and a poor node could lead to more severe impact on the network capacity. Thus, its performance improvement is not as high as other schemes. Both the centralized algorithm and the distributed algorithm obtain consistently higher data rate and lower delay than scheme I under all mean values, as our schemes better alleviate the constraints due to the heterogeneity of the nodes.

7.2 Impact of the Variance of Antenna Array Size

The variance of antenna array size reflects the degree of heterogeneity of nodes in the network. The larger the variance is, the greater the variety of antenna array size is and the portion of poor nodes may become higher. As shown in Fig. 4, when the variance is 0, which is the homogeneous case, the distributed algorithm and scheme I have very close performance. The slightly higher data rate achieved by the distributed algorithm is due to the use of pre-coding. When the variance increases to 0.5, the total rate of the network increases taking advantage of the receiver nodes with larger antenna arrays. The performances of all the distributed algorithms start to decrease when the variance increases beyond 0.5. As the lowest antenna number in the neighborhood is used to constrain the total allowable number of transmission streams in scheme I, its performance degrades faster than scheme II. Our distributed scheme achieves 23% higher rate than scheme I when variance equals 1. As the variance increases further, the performance of scheme I is constrained more by the bottleneck effect at receiver nodes, so the gain of our distributed algorithm constantly increases, achieving up to 36% higher rate and 12% lower delay. This demonstrates that by differentiating between poor nodes and rich nodes and adaptively scheduling transmissions in the network based on the number of antennas at the receiver nodes and channel conditions, our distributed algorithm can effectively alleviate the impact of node heterogeneity and channel variations to achieve better performance. Without being limited by the poorest receiver in the neighborhood, the performance of scheme II reduces slower than other schemes when the variance is extremely high. However, the chance of having extremely high heterogeneity in the network is low, and all the schemes exploiting many-to-many transmissions still achieve significantly higher data rate and lower delay compared to scheme II at the highest variance studied.

7.3 Impact of LOS Component

As described in section 3, the degree-of-freedom of a MIMO channel not only depends on the antenna array size of the end nodes, but also the channel condition of the link. When a LOS component exists, it can impact the correlation between the spatial channels of a link and possibly decrease the degree-of-freedom of the MIMO channel. The impact of the LOS component is described by the K-factor, whose value is generally in the range of $0 \sim 20dB$ in practice. In Fig. 5, the performance of the algorithms under different values of K-factor is studied. Nodes are all equipped with antenna array of size 4. With the increased value of K-factor, i.e., the increase of the strength of the LOS component, the degree-of-freedom of MIMO channels all over the network tends to decrease, which results in fewer orthogonal spatial channels over a link and therefore the degradation in rate and delay for all the schemes. As scheme II mainly relies on the degree of freedom thus the multiplexing gain of a single-link, its performance is impacted most significantly by a stronger LOS component thus lower degree of freedom of the channel, with 70% degradation in data rate and 40% increase in delay. The other algorithms can take advantage of multiuser diversity to schedule streams opportunistically, so the impact

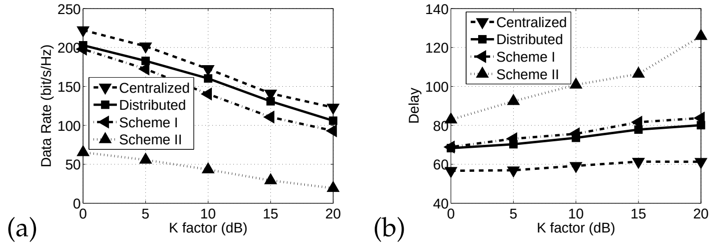

Fig. 5. Impact of LOS component: (a) data rate; (b) delay.

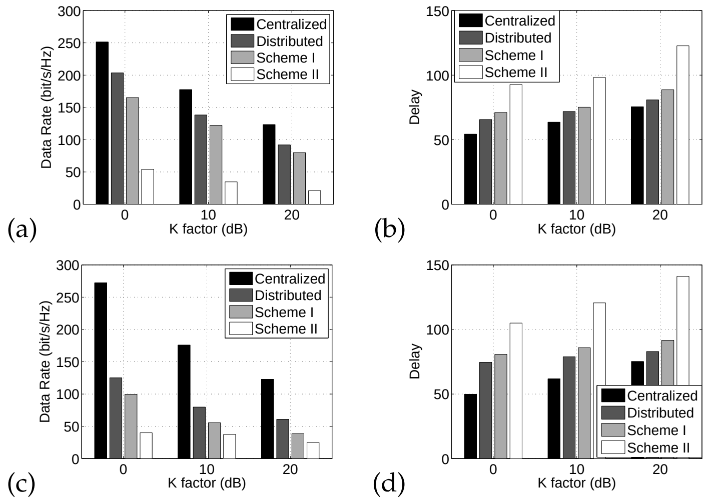

Fig. 6. Impact of LOS component and variance of antenna array size. With variance = 1: (a) data rate; (b) delay. With variance = 2: (c) data rate; (d) delay.

of LOS component on a single link is mitigated with the support of concurrent transmissions among multiple node pairs. Thus the degradations of data rate and delay of scheme I reduce to 55% and 18% respectively. Taking consideration of the impact of the LOS component and channel degree of freedom constraint, our distributed scheduling algorithm adjusts the number of streams accordingly to avoid transmission failure, and obtains up to 18% improvement in data rate and 5% decrease in delay compared to scheme I.

In practice, the degree-of-freedom of a MIMO channel is concurrently impacted by the LOS component and the antenna array size. In Fig. 6 (a) and (b), we present the performance of the algorithms with K-factor of values $0dB$, $10dB$ and $20dB$, and the variance of antenna array size is 1; and in (c) and (d), the variance is changed to 2. The results are consistent with the study of impacts of LOS component and variance of antenna array size independently. With a stronger LOS component at $20dB$ and a larger variance at 2, the distributed algorithm is shown to have more significant performance improvement over both scheme I and II, with 56% higher data rate and 10% lower delay compared to scheme I, and 2.4 times the data rate and 41% lower delay compared to scheme II. The results demonstrate our proposed algorithm is very effective in handling the heterogeneity of network nodes and the variation of channel conditions, and can efficiently exploit the multi-user diversity and antenna selection diversity in a distributed network environment to achieve significant performance gain.

7.4 Impact of Node Density

The impact of node density is shown in Fig. 7. Irrespective of the density, the distributed algorithm has the closest performance to that of the centralized algorithm in terms of both aggregate data rate and normalized delay. As the node

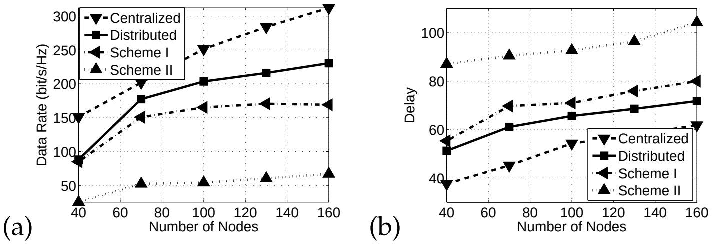

Fig. 7. Impact of node density: (a) data rate; (b) delay.

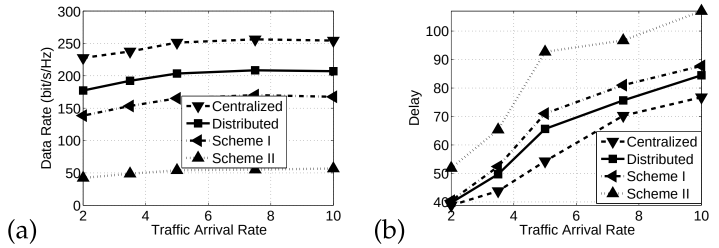

Fig. 8. Impact of traffic arrival rate: (a) data rate; (b) delay.

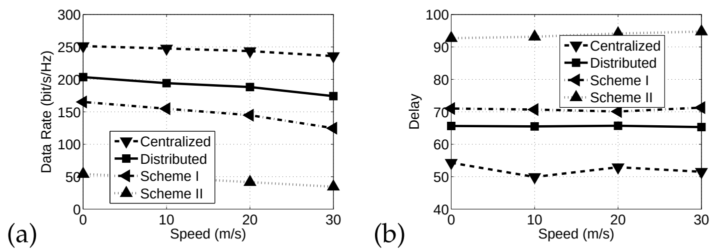

Fig. 9. Impact of mobility: (a) data rate; (b) delay.

density increases, the aggregate data rates of all schemes increase as they can better take advantage of the multiuser diversity. Compared with scheme I, with adaptive selection of transmission strategy based on node types and channel conditions, our distributed algorithm is shown to have up to 36% higher aggregate data rate and 14% lower delay. Compared to scheme II, the distributed algorithm achieves up to 3.6 times the data rate and 31% lower delay. As expected, with only single-user to single-user links, scheme II cannot exploit the transmission potential of nodes and has the lowest data rate and the highest delay.

7.5 Impact of the Traffic Arrival Rate

The traffic arrival rate is denoted by the parameter λ, which is the mean value of the number of packets arrived at the queues of each nodes in each TD, with each queue corresponds to a specific receiver. The value of λ impacts the network performance. If the value is too low, the network is not fully utilized for packet transmission and some of the transmission capacity is wasted. In Fig. 8, when $\lambda = 2$, the data rate is relatively low. If the value of λ is too high, the network may be overloaded which results in an excessive queuing delay of packets. As in Fig. 8, the data rate of each many-to-many scheme initially increases with the increase of traffic and keep almost constant beyond $\lambda \geq 5$, as the network throughput is saturated and cannot accommodate more stream transmissions, while the delay of each scheme is observed to increase due to the longer queuing delay. The reference scheme II has the throughput saturated at a lower traffic arrival rate $\lambda = 3.5$. In order to keep the network in a balanced state, we set $\lambda = 5$ as the default setting. We can also see from the figure that the proposed distributed algorithm outperforms the reference schemes under different traffic arrival rates.

7.6 Impact of Mobility

In our simulation, the positions of nodes change according to the random waypoint model [25] with maximum moving speeds varied, and the path loss varies with the change of the distance between nodes. In addition, the small scale fading is modeled by Rayleigh/Rician fading and independent channel coefficients are generated in each TD to simulate the fast fading effect. As in Fig. 9, the aggregate data rate and delay of all the schemes are not significantly impacted by the mobility, as all of them consider the channel conditions which are impacted by network topology changes and take advantage of the mobility to give transmission preference to the nodes that are closer to the receivers. The proposed distributed algorithm significantly outperforms the reference schemes under all the speeds studied, with up to 36% higher aggregate data rate and 10% lower delay compared to scheme I. The result shows that our algorithm is robust to mobility in the network, as it is always able to coordinate the transmissions based on traffic demand and schedule high-quality streams at any topology. This indicates that our scheme will perform well in a mobile ad hoc network with dynamic topology change.

8 Conclusions

It is important and challenging to coordinate transmissions in a heterogeneous MIMO-based distributed system with mobile devices having different number of antennas, in presence of channel dynamics and network topology changes. In this work, we first formulate the problem to maximize the weighted network data rate and propose a centralized scheduling algorithm with a provable approximation ratio as the performance reference. We then propose an effective distributed scheduling algorithm in MIMO-based ad hoc networks by concurrently considering node heterogeneity, impact of channel condition on the degree of freedom and transmission reliability, traffic demand and network load, and taking advantage of multiuser diversity and spatial diversity. Our algorithm adaptively assumes different transmission strategies based on the decoding capacity of receivers to alleviate the bottlenecks caused by nodes with smaller antenna arrays, and avoid transmission failure due to channel degree of freedom constraint. Our scheduling algorithm also exploits both multiplexing and diversity to opportunistically select transmitter nodes and antennas to improve the transmission rate and reliability, while supporting QoS and fairness. Nodes in a neighborhood can cooperate in transmission and form a many-to-many virtual MIMO array. We form a concrete channel model, and apply the channel model in our algorithm design to efficiently optimize network performance. The performance results demonstrate that our proposed scheduling algorithm is very efficient in coordinating transmissions in a MIMO-based ad hoc network, achieving up to 3.6 times the data rate and reducing the transmission delay up to 31% compared with the scheme of selecting only one user pair at a time as often used in conventional MIMO schemes. Compared with the scheme not considering node heterogeneity and channel constraint, our scheduling algorithm can achieve about 36% higher data rate and 14% lower delay.

Acknowledgements

Xin Wang's research was supported by the US National Science Foundation under grant numbers CNS-0751121 and CNS-0628093. Yuanyuan Yang's research was supported by the US National Science Foundation under grant number ECCS-0801438, and US Army Research Office under grant number W911NF-09-1-0154. This work was done when Shan Chu was with the Department of Electrical and Computer Engineering, Stony Brook University.

References

[1] 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.

[2] J-S. Park, A. Nandan, M. Gerla, and H. Lee. SPACE-MAC: Enabling spatial-reuse using MIMO channel-aware MAC. In *Proc. IEEE ICC 2005*, May 2005.

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

[4] B. Hamdaoui and K.G. Shin. Characterization and analysis of multi-hop wireless mimo network throughput. In *Proc. ACM MobiHoc '07*, pages 120–129, 2007.

[5] J. Liu, Y.T. Hou, Y. Shi, and H. Sherali. Cross-layer optimization for mimo-based wireless ad hoc networks: Routing, power allocation, and bandwidth allocation. *IEEE J. Select. Areas of Commun.*, (6):913–926, August 2008.

[6] K. Sundaresan and R. Sivakumar. A unified MAC layer framework for ad-hoc networks with smart antennas. In *Proc. ACM MobiHoc 2004*, pages 244–255, May 2004.

[7] 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.

[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] P. Casari, M. Levorato, and M. Zorzi. Mac/phy cross-layer design of MIMO ad hoc networks with layered multiuser detection. *IEEE Trans. on Wireless Communications*, (11):4596 – 4607, November 2008.

[10] 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.

[11] J. Liu, Y. Shi, and Y.T. Hou. A tractable and accurate cross-layer model for multi-hop MIMO networks. In *Proc. IEEE INFOCOM '10*, March 2010.

[12] Y. Shi, J. Liu, C. Jiang, C. Gao, and Y. T. Hou. An optimal MIMO link model for multi-hop wireless networks. In *Proc. IEEE INFOCOM '11*, April 2011.

[13] W. Ajib and D. Haccoun. An overview of scheduling algorithms in MIMO-based fourth-generation wireless systems. *IEEE Network*, (5):43–48, Sept.-Oct. 2005.

[14] K. Tan, H. Liu, J. Fang, W. Wang, J. Zhang, M. Chen, and G. M. Voelker. SAM: enabling practical spatial multiple access in wireless LAN. In *Proc. ACM MobiCom '09*, pages 49–60, September 2009.

[15] S. Gollakota, S. D. Perli, and D. Katabi. Interference alignment and cancellation. In *Proc. ACM SIGCOMM 2009*, pages 159–170, 2009.

[16] E. Aryafar, N. Anand, T. Salonidis, and E. W. Knightly. Design and experimental evaluation of multi-user beamforming in wireless LANs. In *Proc. IEEE MOBICOM '10*, 2010.

[17] K. Lin, S. Gollakota, and D. Katabi. Random access heterogeneous mimo networks. In *Proc. ACM SIGCOMM 2011*, 2011.

[18] R. Srinivasan, D. Blough, and P. Santi. Optimal one-shot stream scheduling for mimo links in a single collision domain. In *Proc. IEEE SECON 2009*, pages 700–708, 2009.

[19] S. Chu and X. Wang. Adaptive and distributed scheduling in heterogeneous MIMO-based ad hoc networks. In *Proc. IEEE MASS 2009*, pages 217–226, Oct 2009.

[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] S. Chu and X. Wang. Opportunistic and cooperative spatial multiplexing in MIMO ad hoc networks. *IEEE Trans. on Networking*, (5):1610–1623, Oct. 2010.

[23] R. M. de Moraes, H. R. Sadjadpour, and J. J. Garcia-Luna-Aceves. Many-to-many communication: A new approach for collaboration in MANETs. In *Proc. IEEE INFOCOM 2007*, pages 1829–1837, May 2007.

[24] R. de Lacerda, L. S. Cardoso, R. Knopp, D. Gesbert, and M. Debbah. EMOS platform: Real-time capacity estimation of MIMO channels in the UMTS-TDD band. In *Proc. ISWCS 2007*, pages 782–786.

[25] W. Navidi and T. Camp. Stationary distributions for the random waypoint mobility model. *IEEE Trans. on Mobile Computing*.

Shan Chu

received the B.S. and M.S. degrees in telecommunications engineering from Zhejiang University, Hangzhou, China, and the Ph.D. degree in electrical engineering from Stony Brook University, Stony Brook, NY. She is now with Motorola Solutions, Holtsville, NY. Her current research interests include MIMO and cooperative communications, ad hoc networks and cross-layer design.

Xin Wang

received the BS and MS degrees in telecommunications engineering and wireless communications engineering from Beijing University of Posts and Telecommunications, Beijing, China, and the PhD degree in electrical and computer engineering from Columbia University, New York, New York. She is currently an associate professor in the Department of Electrical and Computer Engineering of the State University of New York at Stony Brook, New York. Before joining Stony Brook, she was a Member of Technical Staff in the area of mobile and wireless networking at Bell Labs Research, Lucent Technologies, New Jersey, and an assistant professor in the Department of Computer Science and Engineering of the State University of New York at Buffalo. Her research interests include analysis and architecture design in wireless networks and communications, mobile and distributed computing, infrastructure design and performance enhancement across network layers, applications and heterogeneous networks, network and mobility management, QoS, signaling and control, as well as support for advanced network services and applications. She is a member of the IEEE.

Yuanyuan Yang

received the BEng and MS degrees in computer science and engineering from Tsinghua University, Beijing, China, and the MSE and PhD degrees in computer science from Johns Hopkins University, Baltimore, Maryland. She is a professor of computer engineering and computer science at Stony Brook University, New York, and the director of Communications and Devices Division at New York State Center of Excellence in Wireless and Information Technology. Her research interests include wireless networks, optical networks, high-speed networks, and parallel and distributed computing systems. Her research has been supported by the US National Science Foundation (NSF) and the US Army Research Office. She has published more than 260 papers in major journals and refereed conference proceedings and holds seven US patents in these areas. She is currently an associate editor for the IEEE Transactions on Computers and a subject area editor for the Journal of Parallel and Distributed Computing. She has served as an associate editor for the IEEE Transactions on Parallel and Distributed Systems. She has served as a general chair, program chair, or vice chair for several major conferences and a program committee member for numerous conferences. She is a fellow of the IEEE and the IEEE Computer Society. More information about her and her research can be found at http://www.ece.sunysb.edu/~yang.