Summer 2014

Cross-layer schemes for performance optimization in wireless networks

Quanmin Ye

Department: Computer Science

CROSS-LAYER SCHEMES FOR PERFORMANCE OPTIMIZATION IN

WIRELESS NETWORKS

by

QUANMIN YE

A DISSERTATION

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

COMPUTER SCIENCE

2014

Approved by

Dr. Maggie Cheng
Dr. Frank Liu
Dr. Fikret Ercal
Dr. Wei Jiang
Dr. Abhijit Gosavi

PUBLICATION DISSERTATION OPTION

This dissertation contains the following five articles:

Paper I: Maggie X. Cheng, Quanmin Ye, Lin Cai, "Cross-Layer Schemes for Reducing Delay in Multihop Wireless Networks", IEEE Transactions on Wireless Communications, 2013, 12(2), pp. 928-937.

Paper II: Maggie X. Cheng, Quanmin Ye, "Transmission Scheduling Based on a New Conflict Graph Model for Multicast in Multihop Wireless Networks", IEEE Globecom 2012.

Paper III: Maggie X. Cheng, Quanmin Ye, X. Cheng, "Simultaneous Routing and Multiplexing in Ad Hoc Networks with MIMO Links", IEEE ICC 2013.

Paper IV: Maggie X. Cheng, Quanmin Ye, "A Combinatorial Solution for Scheduling Spatial Multiplexing in MIMO-Based Ad Hoc Networks", IEEE Globecom 2012, MIMOCR workshop.

Paper V: Maggie X. Cheng, Quanmin Ye, Lin Cai, "Rate-Adaptive Concurrent Transmission Scheduling in 60-GHz mm-Wave WPANs", submitted to IEEE Transaction on Wireless Communications.

ABSTRACT

Wireless networks are undergoing rapid progress and inspiring numerous applications. As the application of wireless networks becomes broader, they are expected to not only provide ubiquitous connectivity, but also support end users with certain service guarantees.

End-to-end delay is an important Quality of Service (QoS) metric in multihop wireless networks. This dissertation addresses how to minimize end-to-end delay through joint optimization of network layer routing and link layer scheduling. Two cross-layer schemes, a loosely coupled cross-layer scheme and a tightly coupled cross-layer scheme, are proposed. The two cross-layer schemes involve interference modeling in multihop wireless networks with omnidirectional antenna. In addition, based on the interference model, multicast schedules are optimized to minimize the total end-to-end delay.

Throughput is another important QoS metric in wireless networks. This dissertation addresses how to leverage the spatial multiplexing function of MIMO links to improve wireless network throughput. Wireless interference modeling of a half-duplex MIMO node is presented. Based on the interference model, routing, spatial multiplexing, and scheduling are jointly considered in one optimization model. The throughput optimization problem is first addressed in constant bit rate networks and then in variable bit rate networks. In a variable data rate network, transmitters can use adaptive coding and modulation schemes to change their data rates so that the data rates are supported by the Signal to Noise and Interference Ratio (SINR). The problem of achieving maximum throughput in a millimeter-wave wireless personal area network is studied.

ACKNOWLEDGEMENTS

I would like to express sincere gratitude to my advisor, Dr. Maggie Cheng, for her guidance, deep knowledge, patience, and encouragement during my Ph.D. study. It is a great experience for me to work with her.

I also would like to thank the members of my advisory committee, Dr. Frank Liu, Dr. Wei Jiang, Dr. Fikret Ercal and Dr. Abhijit Gosavi for their valuable time and their advice to improve the quality of the dissertation.

Thanks to the students of my laboratory, Xuan Gong and Yi Ling for numerous instructive discussions and their kind support.

Finally, I would like to thank my parents, my wife, Danping Qiu, my daughter, Michelle Ye for their infinite love, care and encouragement. With their help, I have overcome the difficulties in my study and life and lived happily in the beautiful town of Rolla, Missouri.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

PAPER IV

PAPER V

LIST OF TABLES

SECTION

1. INTRODUCTION

1.1 WIRELESS NETWORKS

Wireless networks is a type of computer network in which network nodes connect with each other using wireless data connections. Generally, wireless networks are implemented and administered using radio communication. There are many types of wireless networks, such as wireless mesh network, wireless sensor network and so on.

Wireless mesh network is a network topology in which each node relays data for the network. All nodes cooperate in the distribution of data in the network. Unlike traditional WLAN, nodes in wireless mesh network can communicate with each other without access points.

Wireless sensor network is a wireless network, consisting of spatially distributed autonomous sensors. After the initial deployment, sensors start to monitor physical or environmental conditions, for instance, sensor nodes are deployed in several cities to monitor the concentration of dangerous gases for citizens or installed in a forest to detect when a fire has started.

With the increasing popularity of wireless mesh networks and sensor networks, multihop wireless networking technology is expected to not only provide multihop connectivity in locations where wired networks cannot reach, but also to support multimedia applications with stringent quality of service (QoS) requirement. End-to-end delay and throughput are two of the major QoS metrics in wireless networks. This dissertation focus on how to minimize the end-to-end delay and how to maxize the throughput in wireless networks.

1.2 ORGANIZATION OF THE DISSERTATION

In this dissertation, cross-layer schemes for performance optimization are proposed. The proposed schemes are superior to the existing schemes. This dissertation is presented in the form of five chapters as outlined in Figure 1.1. The first two papers focus on how to minimize the end-to-end delay in multihop wireless networks, the last three paper consider how to maximize the throughput in wireless networks.

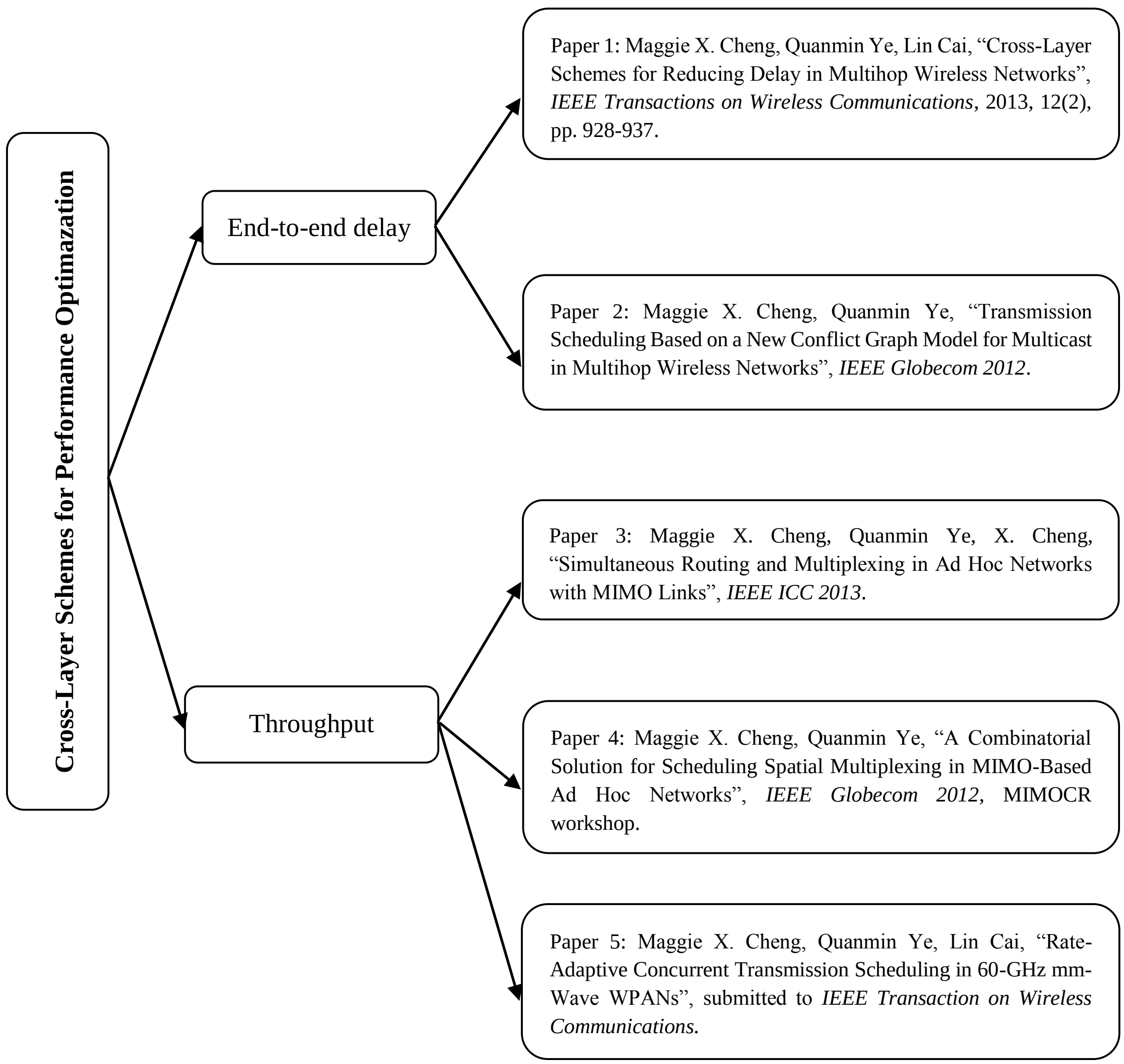


Figure 1.1. Outline of the dissertation

The first paper addresses how to minimize end-to-end delay jointly through optimizing routing and link layer scheduling. Two cross-layer schemes, a loosely coupled cross-layer scheme and a tightly coupled cross-layer scheme are proposed. In the loosely coupled cross-layer scheme, routing is computed first and then the information of routing is used for link layer scheduling; in the tightly coupled scheme, routing and link scheduling are solved in one optimization model. The two cross-layer schemes involve interference modeling in multihop wireless networks with omnidirectional antenna. A sufficient condition on conflict-free transmission is established, which can be transformed to polynomial-sized linear constraints, and a linear program based on the sufficient condition is developed.

The second paper addresses when the multicast tree is given how to schedule wireless nodes for transmission so that network delay is minimized. Firstly the conflict relation among wireless transmissions in a conflict graph is considered, and then a transmission schedule based on an Integer Linear Programming (ILP) model is computed. Since solving ILP problem is NP-hard, a heuristic is designed to solve the ILP problem. The resulting schedule is conflict-free, which is guaranteed by the feasibility of the ILP model.

The third paper addresses how to leverage the spatial multiplexing function of MIMO links to improve wireless network throughput. Wireless interference modeling of a half-duplex MIMO node is presented, based on which, routing, spatial multiplexing and scheduling are jointly considered in one optimization model. A linear program-based algorithm is proposed for the joint optimization.

The fourth paper uses an optimization framework to jointly consider MIMO link spatial multiplexing and scheduling while the routing information is given. A linear program-based algorithm is proposed, and simulation results show it is advantageous over the spatial multiplexing scheme without joint design of scheduling.

Finally, in the fifth paper, the scheduling problem in a millimeterwave wireless personal area network is considered in which users can use adaptive coding and modulation schemes to change their data rates. The scheduling problem is to map transmissions to time slots so that the total throughput is maximized. Discretizing data rate into several distinct levels supported by the PHY layer is proposed, and then use a linear programming model to find the highest rate level a flow can achieve. The same model is extended to consider a mixture of omni-directional antennas and directional antennas with heterogeneous transmitting power.

1.3 CONTRIBUTIONS OF THE DISSERTATION

In the first paper, the main contributions is that two cross-layer design schemes are proposed, and optimization models are established when the impact of wireless interference is considered. The proposed routing and scheduling schemes can outperform their counterparts in each layer, and the integrated cross-layer schemes are superior to the combination of the existing routing and scheduling schemes.

The contributions of second paper is that a node-based conflict graph model is proposed, in which two nodes are considered conflicting if and only if at least one of the receivers of one transmitter is in the interference range of the other transmitter. The conflict graph model dynamically changes with traffic, and accurately captures the conflict relation between nodes. Based on the conflict model, a linear programming model is built to compute the schedule of relay nodes. The proposed scheduling scheme achieves significant delay reduction from the widely used FCFS model.

In the third paper, the major contributions include proposing to jointly consider routing, spatial multiplexing and scheduling at the same time. A linear programming based algorithm is proposed that includes three design problems in one optimization framework.

The contributions of the fourth paper include jointly considering spatial multiplexing and temporal multiplexing in one optimization framework. A linear programming based algorithm is proposed. Simulation results verified the advantage of using such a joint optimization approach.

Finally, the fifth paper focuses on scheduling concurrent transmissions in a variable data rate WPAN. Linear programming-based algorithms for maximum throughput with fairness consideration have been proposed. The proposed schemes showed significant improvement over TDMA and earlier work with concurrent transmissions.

PAPER

I. CROSS-LAYER SCHEMES FOR REDUCING DELAY IN MULTIHOP WIRELESS NETWORKS

Maggie Cheng, Quanmin Ye and Lin Cai

Abstract — end-to-end delay is an important qos metric in multihop wireless networks such as sensor networks and mesh networks. End-to-end delay is defined as the total time it takes for a single packet to reach the destination. It is a result of many factors including the length of the route and the interference level along the path. In this paper we address how to minimize end-to-end delay jointly through optimizing routing and link layer scheduling. We present two cross-layer schemes, a loosely coupled cross-layer scheme and a tightly coupled cross-layer scheme. In the loosely coupled cross-layer scheme, routing is computed first and then the information of routing is used for link layer scheduling; in the tightly coupled scheme, routing and link scheduling are solved in one optimization model. The two cross-layer schemes involve interference modeling in multihop wireless networks with omnidirectional antenna. A sufficient condition on conflict-free transmission is established, which can be transformed to polynomial-sized linear constraints, and a linear program based on the sufficient condition is developed. Through simulation, we show that the proposed routing and scheduling schemes can outperform their counterparts in each layer, and the integrated cross-layer schemes are superior to the combination of the existing routing and scheduling schemes.

I. INTRODUCTION

With the increasing popularity of wireless mesh networks and sensor networks, multihop wireless networking technology is expected to not only provide multihop connectivity in locations where wired networks cannot reach, but also to support multimedia applications with stringent quality of service (QoS) requirement. End-to-end delay is one of the major metrics for quality of service. In this paper endto-end delay refers to the total time it takes for a single packet to reach destination. The user-perceived data transfer speed is a combined effect of both data rate and end-to-end delay. For transferring a small file, the dominating factor is end-to-end delay; for transferring a large file, the dominating factor is data rate. In a typical sensor network, where packets generated by sensors need to be periodically reported to the base station, end-to-end delay plays an important role. This paper aims to address how to achieve minimum end-to-end delay for regular traffic in multihop wireless networks.

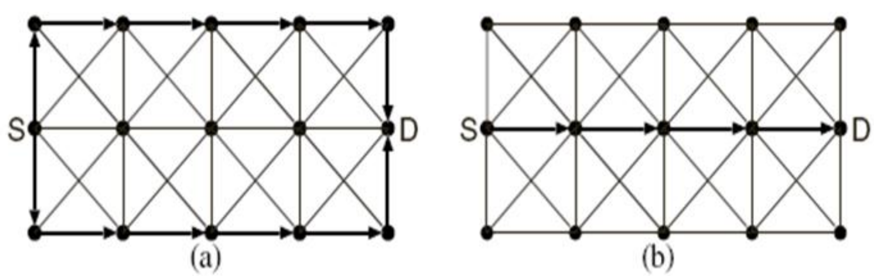


Fig. 1. Latency in maximum throughput routing and minimum delay routing

(a) With maximum throughput routing, latency is 6 slot-time;

(b) With minimum delay routing, latency is 4 slot-time.

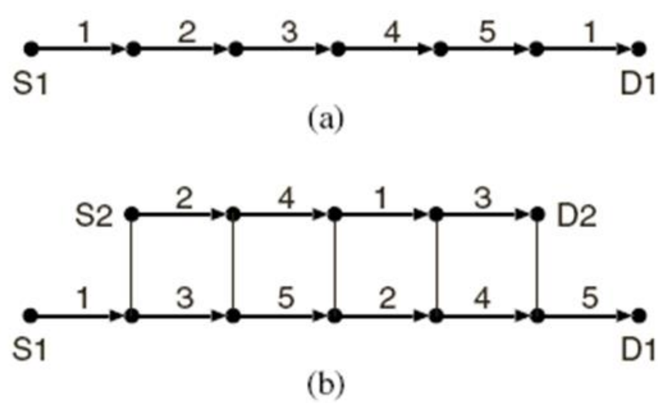


Fig. 2. Latency in one single data flow and in multiple data flows

(a) With a single data flow, latency is 6 slot-time;

(b) When other transmitters are active, the latency becomes 10 slot-time. Numbers on links are slot numbers. There are 5 distinct slot numbers.

In the past, we have seen many reports regarding how to maximize network throughput in multihop wireless networks [1]–[8]. However, the solution that maximizes network throughput often neglects the delay aspect and leads to poor performance in end-to-end delay. For the network in Fig. 1, a maximum throughput routing algorithm would choose (a) since the total throughput from the two paths is twice of that of a single path, and a minimum delay routing algorithm would choose (b) since it is the shortest path and there is no interference from other data flows. Typically, different routes will be selected by the two different routing policies.

In the example shown in Fig. 1, the shortest path happens to have the smallest delay. In this paper, we will demonstrate that it is a misbelief that the shortest path always leads to minimum delay. In fact, end-to-end delay is a result of both the number of hops on the path and the interference level along the path. The shortest path leads to minimum delay only if it is the least interfered path.

Interference works adversely for delay the same way it does for throughput. We define a time slot as the time it takes to transmit a packet over one hop, and the slot duration is assumed constant. In the example shown in Fig. 2, if there is only one data flow from source S1 to destination D1, the end-to-end latency is 6 slots (Fig. 2(a)); however, if there are other transmissions nearby, the end-to-end latency of the same flow can be increased to 10 slots (Fig. 2(b)).

When there are multiple data flows in the network, it is not straightforward to find the optimal paths and transmission schedule that lead to minimum end-to-end delay. In this paper, we propose two cross-layer solutions to reduce end-to-end delay: a loosely coupled scheme and a tightly coupled scheme. The loosely coupled scheme first computes the routing paths that minimize the total interference, then uses a linear programming-based scheduler to decide the time slot assignment such that the end-to-end delay is minimum. The tightly coupled scheme directly uses joint optimization of routing, link rate allocation and slot assignment to achieve minimum delay. Both schemes guarantee that all transmissions are conflict-free in a TDMA context. The main contribution of this paper is that we captured the impact of wireless interference on network delay in the optimization models and the results from these models guarantee conflict-free transmission for TDMA-based multiple access.

The rest of the paper is organized as follows. In Section II, we briefly survey the related work on interference modeling and delay optimization in recent years; in Section III, we present the necessary background for interference modeling in the optimization models, and establish a sufficient condition for conflict-free transmission; in Section IV, we present the loosely coupled scheme and in Section V we present the tightly coupled scheme; in Section VI we show the effectiveness of the proposed schemes through extensive simulations. Finally, Section VII concludes the paper.

II. RELATED WORKS

End-to-end delay is an important performance metric in multihop wireless networks. Delay optimization has been achieved through routing path selection, multiple access control, and sometimes joint design of physical layer and MAC layer.

MAC layer solutions consider the broadcast nature of wireless transmission and reduce the end-to-end delay through transmission scheduling design or transmission probability control. Chatterjee et al. [9] presented when the routing tree is given for a sensor network, how to determine the time slot of each node such that the maximum latency to send a packet from a node to the sink is minimized. Chaporkar et al. [10] addressed the MAC layer multicast problem as an instance of the stochastic shortest path problem and developed an optimal transmission strategy for minimum delay multicasting. Sarkar et al. [11] addressed the energy-delay tradeoff problem and formulated the problem as a constrained optimization problem to minimize the energy consumption while satisfying the constraint on average packet delay. The optimization problem was solved by using dynamic programming formulation, from which a closed form expression for the optimal sleep duration is derived. Pereira et al. [12] addressed delay optimization problem for a random access MAC protocol. They presented an accurate analytical model to derive the optimal transmission probability of each mobile node that minimizes the delay.

Network layer solutions consider delay as one of the constraints while computing the routing path. In [13], Sivrikaya et al. presented an algorithm to compute the minimum-delay path for networks with STDMA. Wan et al. [14] presented approximation algorithms for minimum latency aggregation in sensor networks. The minimum latency problem is to compute an aggregation tree for sensor nodes so that the makespan of the aggregation schedule is minimized. The authors proved that the problem is NP-hard and proposed efficient approximation algorithms. Li et al. [15] studied how to select the routing path with the minimum end-to-end delay in multi-radio wireless mesh networks and developed routing protocols for both single-channel and multi-channel wireless mesh networks. Alzahrani and Woodward [16] proposed a localized QoS routing algorithm by using statistics collected locally and avoided the overhead incurred in most global routing schemes.

Network performance is inherently related to parameters of multiple layers, and strict constraint on one performance aspect at one particular layer is often satisfied at the cost of other layers. Cross-layer joint design and optimization have become a solution to achieve the overall optimal network performance. Cui et al. [17] considered joint design of network layer, MAC layer and physical layer. They optimized the routing flow, TDMA slot assignment and MQAM modulation rate and power on each link to achieve minimum energy and to minimize worst-case packet delay. The cross-layer optimization problems are approximated by convex optimization problems and efficiently solved. Xia et al. [18] used a fuzzy logic system in cross-layer design and considered physical layer, data link layer and application layer together. The cross-layer scheme determines the parameters for adaptive modulation and coding, transmission power, retransmission times, and rate control for packet transmission, and is proved to be effective in improving QoS and energy efficiency. Pakdehi et al. [19] introduced cross layer design between the MAC layer and the physical layer. In this paper, a constrained optimization problem is formulated to optimize the overall system throughput while preserving packet average delay time. Xiao et al. [20] investigated joint design of network-coding and channel-coding. The paper realized the optimal delay performance through the tradeoff design between the network layer and the physical layer. Wang and Shroff [21] addressed the design of network codes and associated flow in network coding in a cross-layer design paradigm, and showed that with a new flow-based characterization of pairwise intersession network coding, a joint optimal scheduling and rate-control algorithm can be implemented distributively. Ukil [22] proposed a cross-layer framework for WiMAX networks to optimize the system performance as well as maintaining the end-toend QoS of individual users, and presented the cross-layer resource allocation and scheduling scheme in the WiMAX system. Vosoughi et al. [23] considered joint design of the MAC layer and the physical layer, and achieved maximum throughput through adapting the power allocation between the source node and relay nodes under the total transmission power constraint.

For interference modeling, the most related work includes [1]–[5]. Jain et al. [1] first used conflict graphs to model the effect of wireless interference under a simplified protocol model; Qiu et al. [3] continued to use conflict graphs to model interference under the IEEE 802.11 interference model; Further in [4], Qiu et al. proposed a physical interference model, which is based on measured interference rather than the distance between nodes. In addition to interference modeling, Padhye et al. [6] focused on the estimation of interference and studied the effect of interference on aggregated network throughput based on the IEEE 802.11 model. In our previous work [24], we proposed a different interference model for directed graphs and used joint routing and link rate control for throughput optimization.

In this paper we propose two cross-layer schemes to improve delay performance. Both schemes involve deterministic combinatorial optimization on route selection and slot assignment. Delay optimization is achieved by jointly considering the effect of path length and interference along the path; Interference is factored in the optimization models for both route selection and slot assignment, which is different from previous works that considered joint optimization of energy or throughput along with delay. Moreover, the accurate endto-end delay modeling allows direct optimization on delay whereas previous works only indirectly or probabilistically improve delay performance.

III. THEORY

A. Bounds on Bandwidth Requirement

To ensure that all transmissions are conflict free, it is important that all active links that are mutually conflicting with each other use different time slots. In other words, two links cannot use the same slot if they interfere with each other.

A collision domain is defined as a group of links that are mutually conflicting with each other. In previous work [1], a conflict graph has been used to model the conflicting relation between wireless links. A conflict graph for a wireless network can be built in polynomial time as follows: we use vertices to represent wireless links, and then put an edge between two vertices if the wireless links they represent interfere with each other. To avoid confusion, we use the terms “vertices” and “edges” in the context of the conflict graph, and use the terms "nodes" and "links" in the context of the wireless connectivity graph.

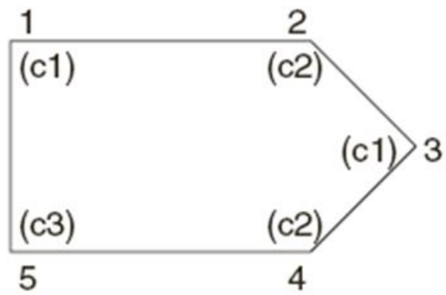


Fig. 3. Lower bound of the total bandwidth needed

$\max\limits_{q:q\in G'}\{\sum_{l\in q} r_l\}$ is only a lower bound of the total bandwidth needed.

To list all collision domains in a network requires to build a conflict graph first and then to find all cliques in the conflict graph. Although to build the conflict graph can be done in polynomial time, to find all cliques in the graph is an NP-hard problem. Moreover, even if we can find all cliques, we only have a lower bound on the total bandwidth needed — Suppose q is a clique in the conflict graph, and wireless link l corresponds to a vertex of the clique, and r_l is the data rate of link l, then $\max\limits_{q}\{\sum_{l\in q} r_l\}$ is the minimum bandwidth needed to ensure all transmissions are conflict-free in a deterministic scheduling scheme. In Fig. 3, the clique number of this graph is only 2, so if we use the approach in [1] for rate allocation, we simply assert $\sum_{l\in q} r_l \leq$ B $\forall$q $\in G'$, then we can only guarantee $r_1 + r_2 \leq$ B, $r_2 + r_3 \leq$ B, $r_3 + r_4 \leq$ B, $r_4 + r_5 \leq$ B, $r_5 + r_1 \leq$ B. If we assign each vertex a data rate of 0.5B, we will satisfy this constraint, but the transmissions still have collisions, because there are always three vertices that cannot transmit at the same time. Let the three vertices be v_1, v_5, and v_4, therefore the total data rate needed by them will be 1.5B, exceeding the total bandwidth limit.

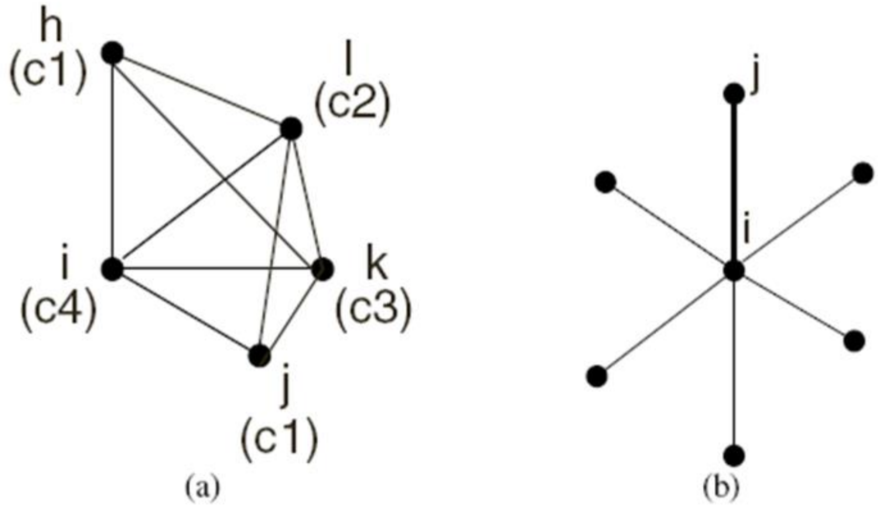


Fig. 4. Bandwidth requirement and proof of sufficiency

(a) Bandwidth requirement (b) Proof of sufficiency.

The obvious upper bound is $\max_i\{r_i + \sum_{j \in N_i} r_j\}$, where i is a vertex of the conflict graph, and N_i are the neighbors of i on the conflict graph. The obvious upper bound tends to be too loose when the neighbors of i are not mutually conflicting with each other. If we use the obvious upper bound, a good bit of bandwidth will be wasted.

In this paper, we propose a tighter upper bound for the bandwidth needed. The upper bound is computable in polynomial time. If we ensure that the upper bound of a rate allocation scheme is at most B, then the rate allocation is feasible, and a deterministic scheduling scheme exists that can schedule all transmissions conflict-free.

Let $G' = (V', E')$ denote the conflict graph. We can use a greedy algorithm to color the conflict graph, which assigns a vertex v the smallest available color not used by v's neighbors. We use V_c' to denote the group of nodes that are assigned color c. The new upper bound UB on the total bandwidth needed can be computed as:

$$\text{UB} = \max_{i:i\in V'}\left\{r_i + \sum_c \max_{j:j\in N_i\cap V_c'} r_j\right\}$$

In contrast, the lower bound defined by the cliques is:

$$\text{LB} = \max_{q\in G'}\left\{\sum_{l:l\in q} r_l\right\}$$

In the conflict graph shown in Fig. 4, vertex i has 4 neighbors and their colors are parenthesized. Vertex i will have different color from any of the neighbors. The total bandwidth needed by this group is $r_i + \max\{r_h, r_j\} + r_l + r_k$, which needs to be kept below total bandwidth B.

If we assert LB $\leq$ B, we only have a necessary condition for conflict-free scheduling; if we assert UB $\leq$ B, we have a sufficient condition for conflict-free scheduling, as stated in the following Theorem. Moreover, to find $\max_{i:i\in V'}\left\{r_i + \sum_c \max_{j:j\in N_i\cap V_c'} r_j\right\}$ we need to compute the maximum value from |V′| vertices in the conflict graph, while to find $\max_{q\in G'}\left\{\sum_{l:l\in q} r_l\right\}$ we have to compute the maximum value from an exponential number of cliques in G′. When the sufficient condition is used as a constraint in the linear program as in (3a), it only expands to a polynomial number of inequalities, whereas the clique condition would expand to an exponential number of inequalities.

Theorem 1: Let $G' = (V', E')$ be the conflict graph for a wireless network. If the following condition is satisfied

$$\max_{i:i\in V'}\left\{r_i + \sum_c \max_{j:j\in N_i\cap V_c'} r_j\right\} \leq B$$

then there exists a conflict-free schedule, in which at any time t, between any pair of vertices i and j in G′ such that $(i, j) \in E'$, there is at most one vertex transmitting.

Proof

The contraposition of the theorem is: If there exists a conflict in slot assignment, then the above condition is not satisfied on some vertex i $\in V'$.

Consider a sequential scheduling algorithm that assigns slots to links in a serial manner. Suppose that there are two links (represented as a pair of vertices on the conflict graph G′) that have a conflict. Let i and j be the first pair that has a conflict. $(i, j) \in E'$, but i and j are assigned to use the same slot t. W.l.o.g. we assume vertex j is assigned the slot before vertex i and we are considering vertex i for the first time. The reason why vertex i has to be assigned to use the same slot as vertex j is that there is no other slot to use. So by the time we consider the slot assignment for vertex i, all the slots have been assigned to the neighbors of i in G′. $\sum_{j:j\in N_i} sl_{j,t} \geq 1\ \forall t$ When this happens, the scheduled transmissions have already used all the bandwidth even without vertex i's transmission, i.e.,

$$\sum_c \max_{j:j\in N_i\cap V'_c} r_j = B$$

If we add the demand from vertex i, then

$$r_i + \sum_c \max_{j:j\in N_i\cap V'_c} r_j > B$$

Since vertex i's data rate r_i is non-zero (Otherwise, i is not active and there is no conflict between i and j). Therefore the constraint is violated. We now can conclude that if the constraint is not violated, then the schedule is conflict-free.■

Remark: Theorem 1 holds for homogeneous networks, in which wireless links have the same capacity B, and therefore the number of time slots a link requires is proportional to its data rate. However in practice wireless links could have different capacities. The formula will need to convert to time scale to effect. Let B_i be the capacity of link i. So r_i/B_i is the percentage of time link i is active. The condition in Theorem 1 can be generalized to the following form for heterogeneous link capacity:

$$\max_{i:i\in V'}\left\{r_i/B_i + \sum_{c} \max_{j:j\in N_i\cap V_c'} r_j/B_j\right\} \leq 1$$

B. The Ratio of the Upper Bound to the Optimal Solution

The conflict graph $G' = (V', E')$ can be generalized to a vertex-weighted graph $G_w' = (V', E', W)$, in which the weight W(i) on vertex i represents data rate, i.e., W(i) = r_i. Let OPT denote the minimum bandwidth requirement to allow conflict-free scheduling. Theorem 1 implies that the upper bound given by UB $= \max_{i:i\in V'}\left\{r_i + \sum_c \max_{j:j\in N_i\cap V_c'} r_j\right\}$ is an upper bound of OPT. Our interest is to find the ratio of the upper bound to the optimal solution.

The problem of finding the accurate bandwidth needed is equivalent to finding the optimal solution of a weighted graph coloring problem. Graph coloring is a way of coloring the vertices of a graph such that no two adjacent vertices share the same color. In the classical graph coloring problem, the objective is to find a proper vertex coloring of the graph that minimizes the total number of colors needed. We call it the cardinality version graph coloring problem. In the weighted version, the objective is to determine a proper vertex coloring of the graph that minimizes the total weight of all color groups, in which the weight of a color group is the maximum weight of all vertices sharing the same color.

The cardinality version is a special case of the weighted version with vertex weight W(v) = 1, ∀ v. The cardinality version graph coloring problem is a well-known NP-hard problem. The weighted version remains NP-hard. Any complexity and approximability result obtained for the weighted version is also applicable to the cardinality version (by simply setting the weight to 1), but the opposite is not true. So far there are a rich collection of results for the cardinality version but there are only some preliminary results for the weighted version, and they are limited to some special graphs [25] and not applicable to the general graphs.

The purpose of this study is to find the ratio of the proved upper bound to the optimal solution. Specifically, we want to know if there is a constant $\boldsymbol{\rho}$ such that

$$\frac{\max_{i:i\in V'}\left\{r_i/B_i + \sum_c \max_{j:j\in N_i\cap V_c'} r_j/B_j\right\}}{OPT} \leq \boldsymbol{\rho}$$

The answer is yes, with $\boldsymbol{\rho} = \Delta(\boldsymbol{G'})$, where $\Delta(\boldsymbol{G'})$ is the maximum node degree in the conflict graph G'.

Theorem 2: The upper bound $\max_{i:i\in V'}\left\{r_i + \sum_c \max_{j:j\in N_i\cap V_c'} r_j\right\}$ is at most $\Delta(\mathrm{G}')\mathrm{OPT}$

Proof

The proof of it requires constructing the worst case scenario by arbitrarily setting the weights. Let M denote a very large number, ε denote a very small number, so that M≫ε. Since to compute the optimal solution is NP-hard, we will use the lower bound of the optimal solution in the ratio analysis. The obvious lower bound is given by $\mathrm{LB} = \max_{q\in G'}\{\sum_{l:l\in q} r_l\}$.

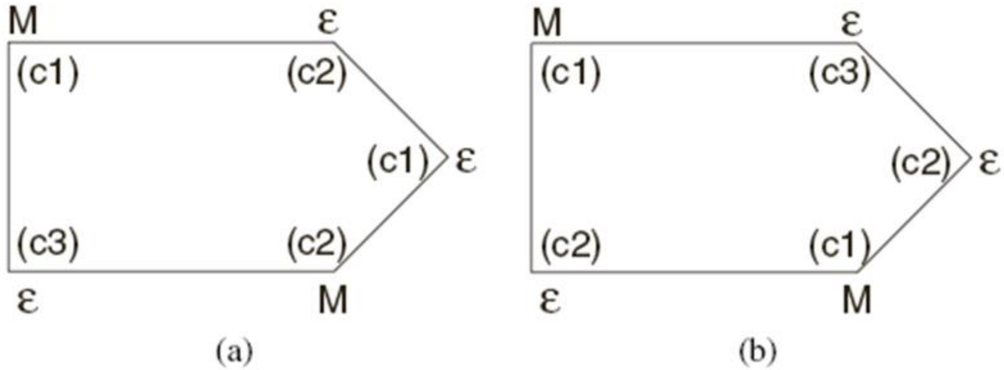


Fig. 5. Greedy solution and optimal solution

(a) A greedy solution; (b) The optimal solution. M and ε are weights, c1, c2, and c3 are colors.

We first look at the example in Fig. 5. The chromatic number 2 of this graph is 3 and the clique number is 2. Suppose the weights of vertices are M and ε as shown in Fig.5. Using a greedy graph coloring algorithm may result in solution (a), from which the upper bound can be obtained: UB = 2M +ε. The optimal solution is (b) with a total weight M + 2ε. The lower bound is M + ε, which does not depend on the color assignment. Note that (a) and (b) use the same number of colors to color the graph, but the total weights are different. The ratio of the upper bound to the optimal solution is 2 in this case.

The graph in Fig. 5 can be generalized to graphs with arbitrarily large chromatic numbers but with small clique numbers. The GrÖtzsch graph is an example of a 4-chromatic graph without a triangle, and the example can be generalized to the Mycielskians. Mycielski's Theorem [26] states that there exist triangle-free graphs with arbitrarily high chromatic number. Since it is triangle-free, the largest clique size is only 2. Now apply a greedy algorithm on such a graph. If a node v is assigned color k > 1, it is because every smaller number from 1 to k-1 has been taken by its neighbors. Select the node with the largest color number and call it node v. If the vertex weights on the neighbors of v that are assigned different colors but are not connected by an edge are all M, and the weight of node v is ε, then the upper bound is $\Delta(G')M + \varepsilon$, while the lower bound is only $M + \varepsilon$. Hence the ratio is $\frac{\Delta(G')M+\varepsilon}{M+\varepsilon} = \Delta(G')$.

If $\frac{UB}{LB} \leq \Delta(G')$, then $\frac{UB}{OPT} \leq \Delta(G')$ since OPT ≥ LB.■

The performance ratio $\Delta(G')$ is obtained for graphs in which weights can be arbitrary. In a real-world network, the weights representing data rates cannot be arbitrary. The limit on data rates will further push down the upper bound closer to the optimal solution. Particularly, when all data rates are equal, the ratio is bounded by $(1+\Delta(G'))/2$, direct from the above analysis by having $M = \varepsilon$. The result is consistent with the cardinality version on triangle-free graphs.

IV. A LOOSELY COUPLED CROSS-LAYER SCHEME

In the following, we present a cross-layer solution that includes routing, link rate control and link transmission scheduling. Our approach is to first find routing paths, and then compute the slot assignment to find the minimum-delay scheduling at the MAC layer. Each subproblem is addressed by solving a separate optimization problem. It is a loosely coupled cross-layer design scheme between the network layer and the MAC layer since the MAC layer only uses the routing information and data rate information to make the scheduling decision, and the network layer uses the interference model from the MAC layer to decide routing paths, but each problem is separately solved. Compared to the second approach (in Section V) which directly computes routing and slot assignment in one optimization problem, the two-phased approach has less computational complexity.

In the following, we will present the network layer solution and the MAC layer solution separately.

A. Minimum Interference Routing (MIR)

Since delay is related to both the hop count of the path, and the interference level that the path is exposed to, we use the total interference along the path as an indicator of delay. This is a better measure of delay than the number of hops, since the same path can experience different amount of delay if the congestion level is different.

We define variable $\mathrm{X}_{ij,f} = 1$ if link (i, j) is on the routing path of flow f. If the flow can be split, $\mathrm{X}_{ij,f}$ can be fractional and hence represent the probability of using link (i, j) for flow f. $\mathrm{R}_{ij,f}$ is the data rate of link (i, j) allocated to flow f. R_f, the source rate of flow f, is an input constant. Each flow is specified by a (source, destination) pair. Since the source and destination of each flow can be arbitrarily selected, it is possible that one node serves as the source for multiple flows to different destinations, so we use $\mathrm{R}_{i,f}$ to denote the source rate of node i for flow f. Since our ultimate goal is to reduce the end-to-end delay of a unit flow, we set R_f the same for all flows. B is the wireless link capacity. I_l roughly gives some indication of the interference level at the location of link l caused by its own transmission and conflicting links' transmissions. I_f is the total interference of all links along the routing path of flow f.

In equality (4c), we quantify I_l as the total non-overlapping "busy time" of link l caused by its own transmissions and nearby transmissions. I_l corresponds to the required bandwidth from link l and its neighbors, given by $r_l + \sum_c \max_{k: k \in N_l \cap V_c'} r_k$. By including the transmission of link l itself, we have implicitly considered the effect of hop count; by including the transmissions on other links in its neighborhood, we have considered the interference that link l received from nearby transmissions.

In the following, we use l to denote the undirected link between node i and node j, and use (i, j) to denote the directed link from i to j. The minimum interference routing (MIR) can be modeled as follows:

Minimize

$$\sum_f I_f \tag{1}$$

Subject to

(1) Flow conservation

$$\sum_{j \in N_i} \left(R_{ij,f} - R_{ji,f}\right) = R_{i,f} \ , \forall i, f \tag{2a}$$

$$R_{i,j} = \sum_f R_{ij,f} \ , \forall link(i,j) \tag{2b}$$

$$R_{i,f} = R_f \ , \forall f, i \ is \ Source(f) \tag{2c}$$

$$R_{i,f} = -R_f \ , \forall f, i \ is \ Sink(f) \tag{2d}$$

$$R_{i,f} = 0 \ , \forall f, i \neq Source(f), Sink(f) \tag{2e}$$

(2) Bandwidth constraint

$$r_l + \sum_c \max_{k: k \in N_l \cap V_c'} r_k \leq B, \forall \ l \in V' \tag{3a}$$

$$r_l = R_{ij} + R_{ji} \ , \forall link \ l \tag{3b}$$

(3) Interference modeling

$$R_{ij,f} = R_f \cdot X_{ij,f} \ , \forall flow \ f \tag{4a}$$

$$I_f = \sum_{link(i,j)} I_l \cdot X_{ij,f} \ , \forall flow \ f \tag{4b}$$

$$I_l = \ r_l + \sum_c \max_{k:k\in N_l \cap V_c'} r_k \ \forall link\ l \qquad (4c)$$

$$0 \leq \ X_{ij,f} \leq 1 \ , 0 \leq \ R_{ij} \leq B \qquad (4d)$$

The above formulation is not linear since both I_l and $\mathrm{X}_{ij,f}$ in constraint (4b) are variables. A close approximation to the above formulation is to minimize the total interference from all links, i.e.,

Minimize

$$\sum_{link\ l} I_l \qquad (5)$$

Subject to All constraints but (4b)

Without constraint (4b) the alternative formulation is linear. Minimizing total interference from all links can indirectly reduce the end-to-end delay. If the flow is allowed to split, then $\mathrm{X}_{ij,f}$ is real-valued, then the above linear program can be solved in polynomial time; If the flow cannot be split, then $\mathrm{X}_{ij,f}$is a 0-1 integer variable, in which case we can relax it to a fractional linear program and then use LP-rounding based scheme to find the routing paths.

The following iterative algorithm can be used to solve the nonlinear program with objective function (1):

1. solve the alternative linear program (5) to get $\mathrm{X}_{ij,f}$ and I_l, calculate I_f using equation (4b), then set objective value $Z_0 = \sum_f I_f$;
2. use I_l as the link weight; for each source, use the shortest path algorithm to find the path with the smallest total interference;
3. update $\mathrm{X}_{ij,f}$, I_l and I_f; set $Z_1 = \sum_f I_f$;
4. if $Z_1 < Z_0$, set $Z_0 = Z_1$; otherwise stop.
5. Repeat steps 2–4 until the maximum number of iterations are reached.

The algorithm will stop either because it is no longer improving or the maximum number of iterations is reached. In the simulation, most examples take 3 ~ 5 iterations to stop. The nonlinear program only requires to solve the linear program for once to get the initial values, and the iterative procedure only involves computing the shortest paths and updating the variables. The overall time-complexity is still dominated by solving the linear program, which takes $O(n^3)$ for n variables. After we solved $X_{ij,f}$, the routing information is supplied to the MAC layer to assist link scheduling.

B. Minimum Delay Link Scheduling (MinDelay)

In this section, we show how to reduce the end-to-end delay by scheduling link transmissions when the routing path is given. When a relay node forwards a packet, there is a mandatory store-and-forward delay and a link scheduling delay that is dependent on the scheduling policy. Link scheduling delay is introduced when the outgoing link uses a time slot that is not immediately after the slot used by the incoming link. If the outgoing link uses slot number v, and the incoming link uses slot number u, the total delay introduced at relay node r is $d_r = v - u$ if $v > u$, or $d_r = v - u + T$ if $v < u$, where T is the total number of distinct slots in a scheduling period (or superframe) in the TDMA context. Fig.8 shows an example. If the schedule is conflict-free, it is guaranteed $v \neq u$. The end-to-end delay for a path is $\sum_r d_r$. From this formula we can see that end-to-end delay is related to both the total number of hops, and the scheduling delay at each relay node. When the routing information is given, the only factor that can be optimized in the MAC layer is the scheduling delay.

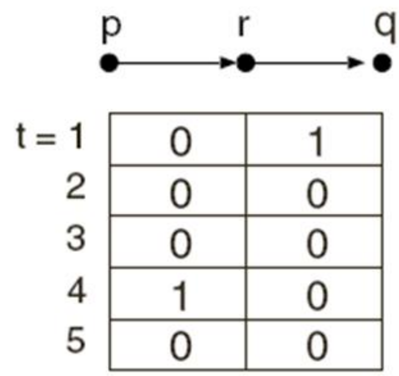


Fig. 6. Total delay at relay node

With T = 5, link (p, r) uses slot 4 and link (r, q) uses slot 1, the total delay at node r is 1-4+5=2. The values in $sl_{pr,t}$ and $sl_{rq,t}$ are shown in picture.

1) An ILP model for Minimum Delay Scheduling: To achieve minimum scheduling delay, we first formulate it as an optimization problem. Since the routing information is given, we use constant $\text{link}_{l,f} = 1$ to indicate link l is on the path for flow f. What we need to solve is the slot assignment for links. We introduce a 0-1 variable $\text{sl}_{l,t}$ for slot assignment. $\text{sl}_{l,t} = 1$ indicates link l uses time slot t. If a link l is shared by multiple data flows, only one flow can use it to transmit at time t. Flows can have different data rates, so we use R_f to denote the data rate of flow f. $\text{sl}_{l,f,t,i} = 1$ indicates link l uses slot t for sending the i^{th} unit of data from flow f. Each unit of data (e.g., a packet) can be transmitted during one slot time. The model presented in the following is a generalization of the MinDelay scheme in [27], which only considers unit flows, whereas this model can be applied to flows of different source rates.

Assume for flow f, relay node r is on the routing path P_f with incoming link l_1 and outgoing link l_2. For the i^{th} unit of data from flow f, the delay at relay node r is denoted by $\text{d}_{f,r,i}$. We define $\text{x}_{f,r,i}$ as a boolean variable: $\text{x}_{f,r,i}$ = 1 when the slot number for the outgoing link is smaller than the slot number for the incoming link; otherwise $x_{f,r,i}$ = 0. The integer linear programming model to minimize the total delay is now formulated as follows:

Minimize

$$\sum_{f}\sum_{r\in P_f}\sum_{i=1}^{R_f} d_{f,r,i} \tag{6}$$

Subject to

$$sl_{l,t} + sl_{l',t} \le 1, \forall (l,l') \in E', \forall t \tag{7a}$$

$$sl_{l,t} = \sum_{f}\sum_{k=1}^{R_f} sl_{l,f,t,k}, \forall l, \forall t \tag{7b}$$

$$\sum_{t=1}^{T} sl_{l,f,t,k} = link_{l,f}, \forall l, \forall f, \forall k \in \{1, \dots, R_f\} \tag{7c}$$

$$d_{f,r,k} = \sum_{t=1}^{T} sl_{l2,f,t,k} \times t - \sum_{t=1}^{T} sl_{l1,f,t,k} \times t$$

$$+x_{f,r,k} \times T, \forall f, \forall r \in P_f, \forall k \in \{1, \dots, R_f\} \tag{7d}$$

$$sl_{l,t} = \{0,1\}, sl_{l,f,t,k} = \{0,1\} \tag{7e}$$

$$0 < d_{f,r,k} < T, x_{f,r,k} \in \{0,1\} \tag{7f}$$

Inequality (7a) requires that for any slot t, if two links l and l′have conflict, at most one of them can use slot t; (7b) indicates that as long as link l is used to transmit some data at time t, then $sl_{l,t}$= 1; otherwise, $sl_{l,t}$ = 0; (7c) indicates that if link l is on the routing path, then each unit of data is needs a slot, therefore the total number of slots assigned to this link for this flow is equal to the total data rate on the link; (7d) quantifies the delay experienced at relay node r for a unit of data.

If the linear program is feasible, then there is a conflict-free schedule. If we use the MIR algorithm to generate routing information, it is guaranteed that the above integer linear program is feasible. The objective function is the total delay. The formula considers the path length and the actual delay at each relay node, which is determined by the local interference level.

2) Computing the slot assignment: To solve the above integer linear program is NP-hard. We first relax it to a real-valued linear program, then use the following rounding algorithm to map real numbers to integers. Solving the linear program takes $O(n^3)$-time, while the rounding part takes O(nlgn)-time for n variables.

a) Sort $sl_{l,t}$ in non-increasing order, set T_h = 0.5;

b) Pick the largest $sl_{l,t}$ in the list, if $sl_{l,t} \geq T_h$, assign $sl_{l,t}$ = 1. Assign $sl_{l',t}$ = 0 for other links l' that are conflicting with l. Assign remaining values appropriately to satisfy flow conservation; If T_h> the largest $sl_{l,t}$ set T_h = the largest $sl_{l,t}$;

c) Repeat step 2) until all variables are rounded to integers.

V. JOINT DESIGN OF ROUTING AND SCHEDULING FOR REDUCING DELAY

Here we present a tightly coupled cross-layer scheme, in which the network layer and the MAC layer solutions are computed in one optimization problem. Solving this optimization problem will get routing paths, link data rates as well as time slot assignment for transmission on each link.

Let R_{ij} be the data rate on link (i, j) (from i to j), and $R_{ij,f}$ be the data rate on link (i, j) allocated to flow f. We use 0-1 variable $X_{ij,f}$ to indicate link (i, j) is used to transmit data from flow f. Decision variable $sl_{ij,t}$ = 1 indicates slot t is used by the directed link (i, j). If i and j are two endpoints of link l, we define $sl_{l,t}$ = $sl_{ij,t}$+ $sl_{ji,t}$. In fact $sl_{l,t}$ is also a

0-1 variable, since (i, j) and (j, i) cannot both use slot t. Subscript l is used to identify an undirected link, while (i, j) is used to denote a directed link from i to j.

Assume flow f passes through relay node r. Relay node r receives flow from incoming link (p, r) and forwards it to outgoing link (r, q), the delay at relay node r is denoted as $d_{f,pr,rq,k}$ for the k^{th} packet. When a relay node forwards a packet, there is a mandatory store-and-forward delay and a variable queueing delay that is dependent on the scheduling policy. The end-to-end delay for a flow is the sum of all delays along the path.

We now can formulate the optimization problem as follows.

Minimize

$$\sum_{f}\sum_{r}\sum_{i=1}^{R_f} d_{f,r,i} \tag{8}$$

Subject to

(1) Flow conservation

$$\sum_{j\in N_i}\left(R_{ij,f} - R_{ji,f}\right) = R_{i,f}\ ,\forall i, f \tag{9a}$$

$$R_{i,f} = R_f\ ,\forall f, i\ is\ Source(f) \tag{9b}$$

$$R_{i,f} = -R_f\ ,\forall f, i\ is\ Sink(f) \tag{9c}$$

$$R_{i,f} = 0\ ,\forall f, i \neq Source(f), Sink(f) \tag{9d}$$

$$R_{ij,f} = R_f \times X_{ij,f}\ ,\forall f, \forall link(i,j) \tag{9e}$$

(2) Delay modeling

$$sl_{l,t} + sl_{l',t} \le 1, \forall (l,l') \in E'\ ,\forall t \tag{10a}$$

$$sl_{l,t} = sl_{ij,t} + sl_{ji,t}\ ,\forall l(i,j), \forall t \tag{10b}$$

$$sl_{ij,t} = \sum_{f}\sum_{k=1}^{R_f} sl_{ij,f,t,k}\ ,\forall link(i,j), \forall t \tag{10c}$$

$$\sum_{t=1}^{T} sl_{ij,f,t,k} = X_{ij,f}\ ,\forall link(i,j), \forall f, \forall k \in \{1, \dots, R_f\} \tag{10d}$$

$$d'_{f,pr,rq,k} = \sum_{t=1}^{T} sl_{rq,f,t,k} \times t - \sum_{t=1}^{T} sl_{pr,f,t,k} \times t$$
$$+x_{f,pr,rq,k} \times T, \forall f, \forall r \neq Source(f) \text{ or } Sink(f),$$
$$\forall \mathrm{p}, \mathrm{q} \in N_r, \forall k \in \{1, \dots, R_f\} \quad (10e)$$
$$d_{f,pr,rq,k} = \max\{d'_{f,pr,rq,k} + (X_{pr,f} + X_{rq,f} - 2)N_\infty, 0\}$$
$$\forall f, \forall r \neq Source(f) \text{ or } Sink(f), \forall \mathrm{p}, \mathrm{q} \in N_r, \forall k \in \{1, \dots, R_f\} \quad (10f)$$
$$d_{f,r,k} = \sum_{p,q \in N_r} d_{f,pr,rq,k}, \forall f, \forall r, \forall k \in \{1, \dots, R_f\} \quad (10g)$$
$$0 \leq d'_{f,pr,rq,k}, d_{f,pr,rq,k}, d_{f,r,k} \leq T, x_{f,pr,rq,k} \in \{0,1\},$$
$$X_{ij,f} = \{0,1\}, sl_{l,t}, sl_{ij,f,t,k} = \{0,1\} \quad (10h)$$

In (10f), N_∞ is a very large positive integer. When link pr or rq is not on the routing path for flow f, at least one of $X_{pr,f}$ and $X_{rq,f}$, is zero, so $X_{pr,f}$ $+X_{rq,f} - 2$ is negative, and $d'_{f,pr,rq,k} + (X_{pr,f} + X_{rq,f} - 2)N_\infty$ will be negative, therefore $d_{f,pr,rq,k}$ will be set to zero.

The above optimization model has integer variables. We first relax them to real numbers, solve the linear program and then use iterative rounding to find the integer solutions. For example, for slot assignment, the rounding process starts from sorting the fractional solutions, and then round up the largest $sl_{l,t}$ to 1, and then set all other links that conflict with this link to be 0, and assign the remaining values appropriately to satisfy flow conservation. Repeat this process until all values are either rounded to 0 or 1. Using a procedure similar to the rounding algorithm in Section IV we can solve all the integer variables.

After solving the ILP problem, the network layer and the MAC layer solution can be retrieved. For example, $X_{ij,f}$= 1 indicates link (i, j) is used on the routing path for flow f; and $sl_{ij,f,t,k}$ = 1 for any k indicates link (i, j) will use slot t to send data from flow f. Routing solution as well as link scheduling solution are all solved in one scheme.

The optimization problem is defined on an space with multiple dimensions. As the optimization space becomes larger, the number of variables increases. The advantage of using a large optimization space is that if it can be solved, the solution will result in shorter delay than the solution obtained from the smaller optimization space; the disadvantage is that solving the problem is less efficient, as we can see from the simulation results in section VI. We conclude that for a small network, we can use the tightly coupled routing and scheduling scheme; but for a large-scale network, it is reasonable to compromise the performance for a faster solution.

It is worth mentioning that the bandwidth constraint is implicitly expressed in inequality (10a):

$$sl_{l,t} + sl_{l',t} \leq 1, \forall (l, l') \in E', \forall t$$

There are T distinct slots (t = 1, 2, ..., T), which is the number of slots in a superframe. T is proportional to the link bandwidth B — If the data rate is represented in the number of *packets*/*cycle*, then B=T. If the linear program is infeasible, it is because constraint (10a) is violated, which indicates there is not enough bandwidth to support the traffic demand.

VI. SIMULATION

In this section, we evaluate the performance of the proposed algorithms through simulation. We first use a separate-layer approach to compare them with the well-known algorithms at each layer, and then evaluate the overall performance of the cross-layer schemes.

A. Separate-layer Comparison

In the loosely coupled schemes, we proposed a routing scheme Minimum Interference Routing (MIR), and a link level scheduling algorithm MinDelay. MinDelay (with objective function (1)) uses the routing information from the network layer for link-level transmission scheduling and can be used with any routing algorithm. MIR as a routing algorithm can work with any MAC layer scheme. We now compare MinDelay with First-Come First-Served (FCFS) at the MAC layer when MIR and Shortest Path Routing (SPR) are used for routing solution respectively; and compare MIR with SPR when FCFS and MinDelay are used for scheduling respectively.

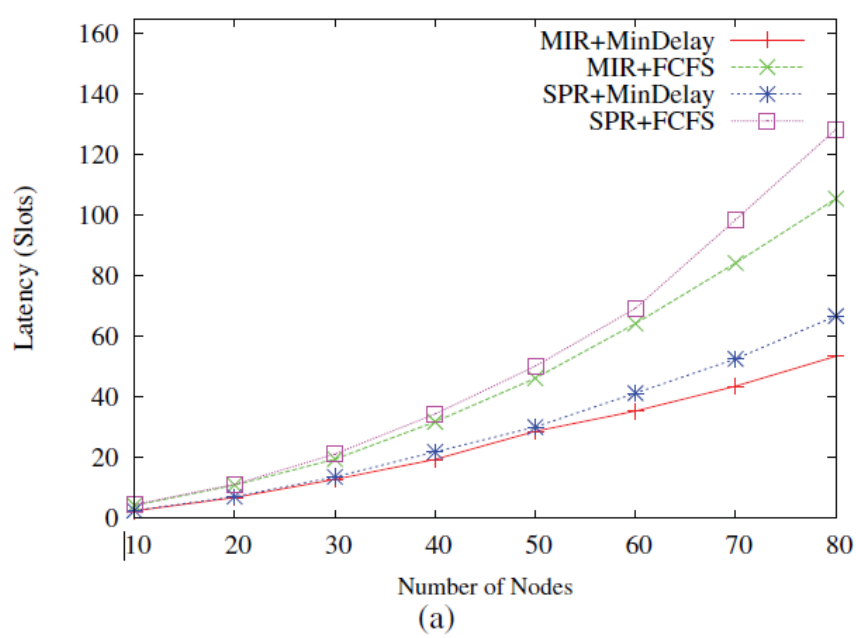
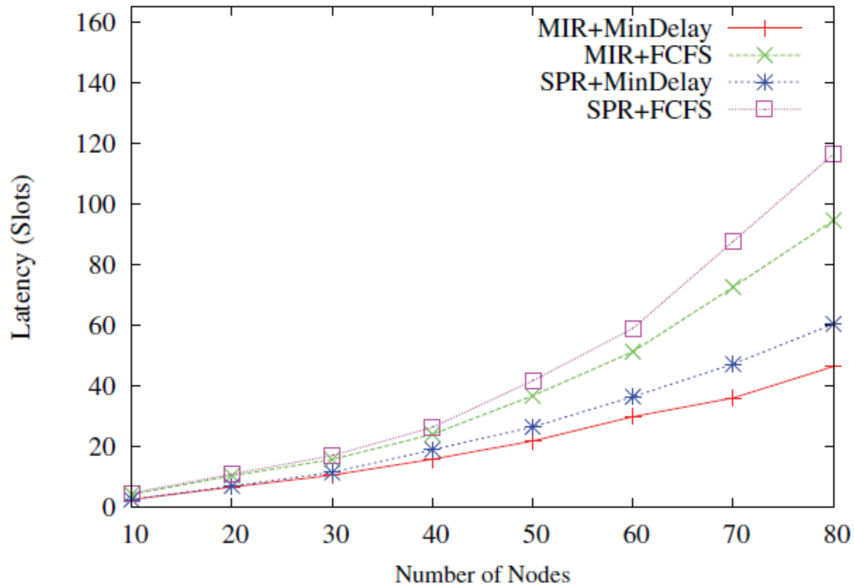


Fig. 7. Separate-layer comparison: MinDelay vs. FCFS, MIR vs. SPR



(b)

Fig. 7. Separate-layer comparison: MinDelay vs. FCFS, MIR vs. SPR (cont.)

In FCFS, a relay node schedules a packet as soon as it arrives; when deciding which slot to use, a relay node chooses the next available slot to transmit the packet if it does not conflict with other scheduled transmissions. FCFS is the most commonly used scheduling policy in practice. In SPR, a router choose the shortest path (in hops) to reach the destination regardless of other transmissions. It is used in practice for its simplicity.

In the simulation study, we use 10–80 nodes deployed in a 150m× 150m square region, with node transmission range 30m. 20% nodes are randomly selected as source nodes. Each source f generates R_f(the given source rate) packets in a TDMA cycle, and through multihop forwarding the packets are delivered to the destination. We observe the total delay of all flows. For each network size, we repeat the simulation for 20 runs with randomly generated network topology and plot the average delay. Fig. 7(a) shows the result when all sources send to a common destination, and (b) shows the result when each source has its own destination.

Simulation results in Fig. 7(a) show that on average when using SPR, MinDelay outperforms FCFS by 36% to 48% in total delay; and when using the proposed routing algorithm MIR, MinDelay outperforms FCFS by 34% to 49%. From this simulation we observed MinDelay has much lower latency than FCFS in all scenarios regardless of which routing algorithm is used. From the comparison of the two routing schemes, we also observed that no matter which MAC layer scheduling scheme is used, MIR always outperforms SPR. The performance gain is between 4% to 20% when MinDelay is used, and is between 3% to 18% when FCFS is used. It is shown that the performance gain is larger when the network size is larger, since there is more room to optimize when the number of hops and interference level increase. Fig. 7(b) shows the same trend with less delay than (a) due to the fact that having separate destination for each flow can alleviate contention near the common destination.

B. Overall Comparison

In the second simulation, we evaluate the two loosely coupled schemes and the tightly coupled scheme on end-to-end delay. Three curves in Fig.8 are for loosely coupled linear scheme (Linear), loosely coupled nonlinear scheme (NonLinear), and the tightly coupled scheme (Joint). Since the tightly coupled algorithm works slowly in large network topology, we use a small network with 10 nodes randomly deployed in a 65m × 65m area. Three sources have been randomly chosen to generate packets. Fig. 8(a) shows total delay when all sources send to one common sink, and (b) shows total delay when each source sends to its own destination. When choosing source and destination nodes, we make sure each source is at least three hops away from its destination in order to show the difference between different routing schemes. The simulation results show that the tightly coupled scheme outperforms the NonLinear loosely coupled schemes by 15% and the Linear loosely coupled scheme by 24% in end-to-end delay (see Fig. 8.). However the time it takes for solving the linear program in the tightly coupled scheme is much longer than in the loosely coupled schemes due to the large number of variables. The running time of the tightly coupled scheme is 7.8 to 10.4 times longer than that of the loosely coupled schemes. Due to the small size of the network, having a different destination does not improve the overall interference level as much as it does for a large network, so (b) is only slightly better than (a) in average end-to-end delay.

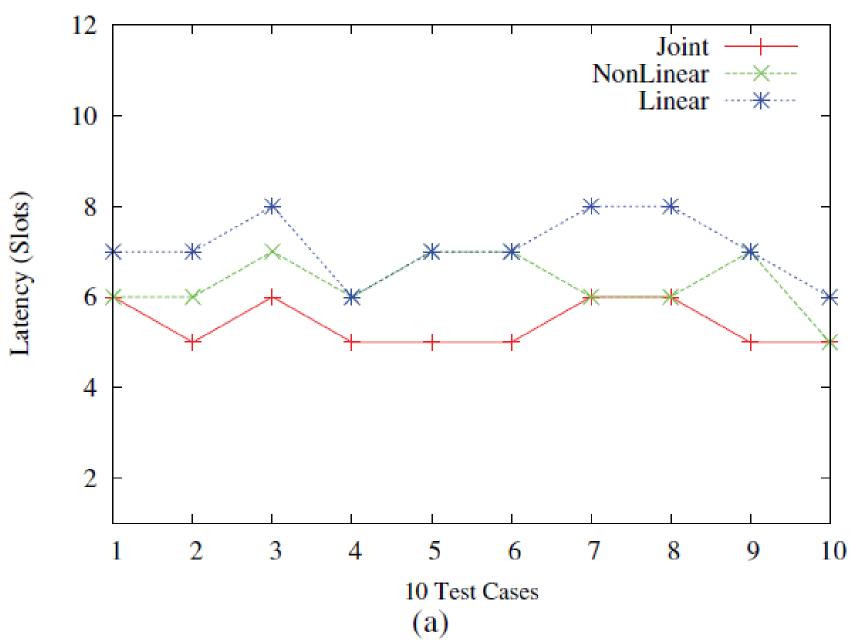


Fig. 8. Cross-layer comparison on a 10-node network

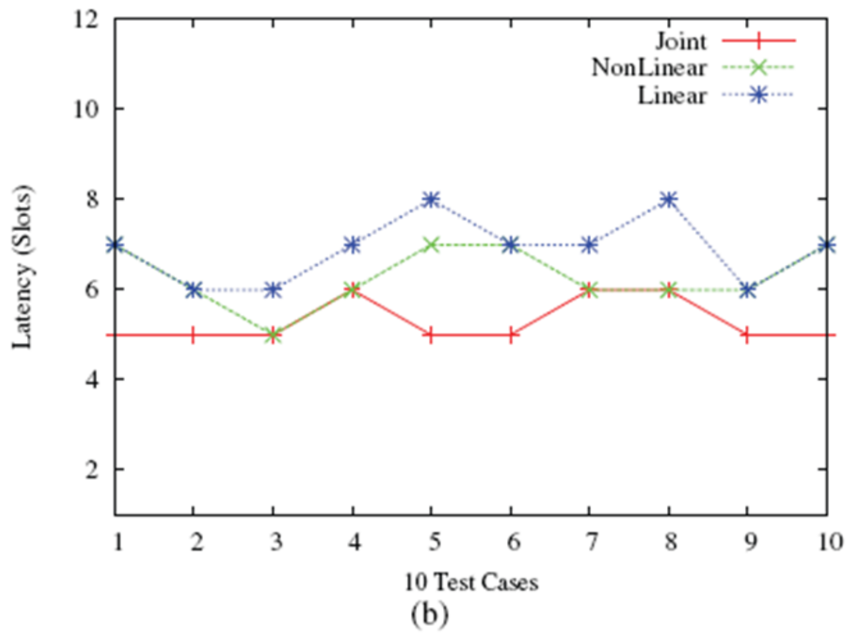


Fig. 8. Cross-layer comparison on a 10-node network (cont.)

Cross-layer comparison on a 10-node network. Loosely coupled schemes vs. the tightly coupled scheme.

Due to the running time inefficiency of the tightly coupled scheme, in practice it is recommended to use the tightly coupled scheme for small networks only. For larger network sizes, the loosely coupled schemes are more suitable. We compare the overall performance of the two loosely coupled schemes with the commonly used schemes in practice, and observe the combined effect of both routing and scheduling. Fig. 9 shows the results for network sizes between 10 to 80 nodes. The deployment of the network is the same as in section VI-A. Fig. 9(a) shows total delay when all sources send to the same destination and Fig. 9(b) shows total delay when each source sends to its own destination. We make sure the number of hops from sources to destinations are at least 3 hops for 10-node networks, at least 7 hops for the 25-node networks, at least 13 hops for the 50-node networks, and at least 17 hops for the 80-node networks. The NonLinear scheme is marginally better than the Linear scheme, and both are substantially better the Shortest Path Routing with FCFS scheduling in all scenarios. In Fig. 9(a), for 10-node networks, NonLinear outperforms SPR+FCFS by 45% and Linear outperforms FCFS+FCFS by 43%; For 80-node networks, NonLinear outperforms SPR+FCFS by 60% and Linear outperforms SPR+FCFS by 51%. When the network size increases, the performance gain is higher. Fig. 9(b) shows less delay than in (a) in all scenarios due to the space separation of different flows.

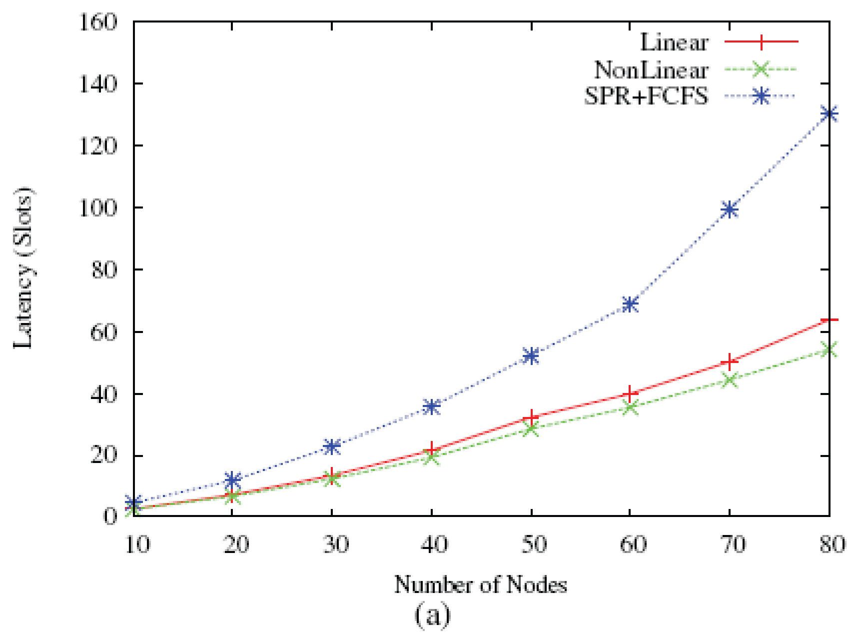
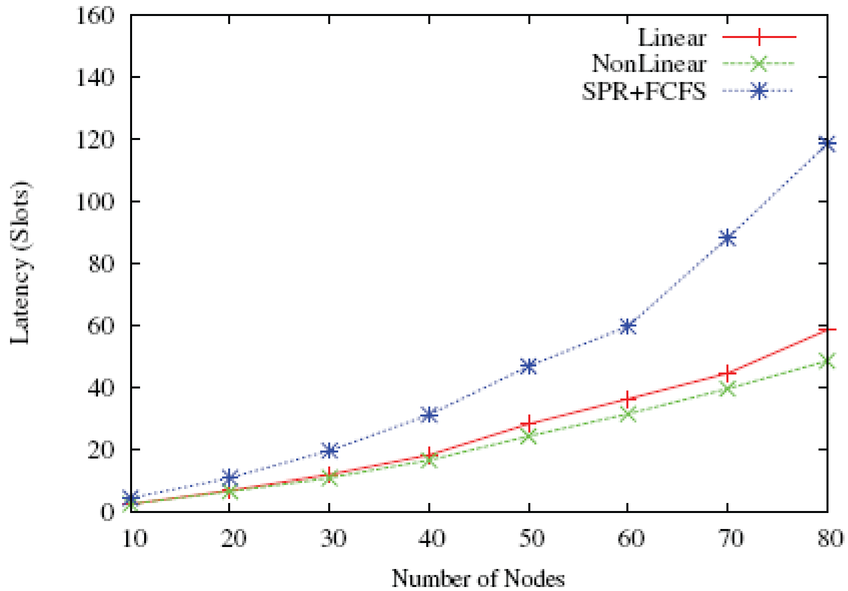


Fig. 9. End-to-end delay comparison for networks of 20 to 80 nodes



(b)

Fig. 9. End-to-end delay comparison for networks of 20 to 80 nodes (cont.)

End-to-end delay comparison for networks of 20 to 80 nodes, the loosely coupled schemes vs. SPR+FCFS.

VII. CONCLUDING REMARKS

In this paper, we have studied how to achieve minimum end-to-end delay in a multihop wireless network. We have presented two cross-layer design schemes, and established optimization models when the impact of wireless interference is considered. A sufficient condition for conflict-free transmission is established, and a linear programming model for minimum interference routing is developed using this condition as a constraint. The simulation results show that model-based optimization did achieve shorter delay than the existing routing and scheduling schemes.

The sufficient condition is used in a global optimization framework in this paper, but it can also be used locally for dynamic scheduling, in which the condition only needs to be satisfied within the 2-hop neighborhood of a link. Moreover, it can also be used for resource reservation and admission control as part of QoS provisioning in wireless networks. We will address this issue in the future work.

REFERENCES

[1] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu, “Impact of interference on multi-hop wireless network performance,” in *Proc. 2003 International Conf. Mobile Comput. Netw*., pp. 66–80.

[2] S. Rangwala, R. Gummadi, R. Govindan, and K. Psounis, “Interferenceaware fair rate control in wireless sensor networks,” in *Proc. 2006 Conf. Appl., Technol., Architectures, Protocols Comput. Commun.*, pp. 63–74.

[3] Y. Li, L. Qiu, Y. Zhang, R. Mahajan, Z. Zhong, G. Deshpande, and E. Rozner, “Effects of interference on wireless mesh networks: pathologies and a preliminary solution,” in *2007 HotNets*.

[4] L. Qiu, Y. Zhang, F. Wang, M. K. Han, and R. Mahajan, “A general model of wireless interference,” in Proc. *2007 ACM International Conf. Mobile Comput. Netw*., pp. 171–182.

[5] Y. Li, L. Qiu, Y. Zhang, R. Mahajan, and E. Rozner, “Predictable performance optimization for wireless networks,” in *Proc. 2008 ACM Conf. Data Commun*., pp. 413–426.

[6] J. Padhye, S. Agarwal, V. N. Padmanabhan, L. Qiu, A. Rao, and B. Zill, “Estimation of link interference in static multi-hop wireless networks,” in *Proc. 2005 ACM SIGCOMM Conf. Internet Measurement*, pp. 28–28.

[7] M. Kodialam and T. Nandagopal, “Characterizing achievable rates in multi-hop wireless networks: the joint routing and scheduling problem,” in Proc. 2003 International Conf. Mobile Comput. Netw., pp. 42–54

[8] M. Alicherry, R. Bhatia, and L. Li, “Joint channel assignment and routing for throughput optimization in multiradio wireless mesh networks,” in *2005 ACM MobiCom*.

[9] P. Chatterjee and N. Das, “A cross-layer distributed TDMA scheduling for data gathering with minimum latency in wireless sensor networks,” in *Proc. 2009 Wireless VITAE*, pp. 813–817.

[10] P. Chaporkar and S. Sarkar, “Minimizing delay in loss-tolerant MAC layer multicast,” in *Proc. 2005 International Symp. on Modeling Optimization Mobile, Ad Hoc, Wireless Netw*., pp. 358–367.

[11] M. Sarkar and R. L. Cruz, “A MAC layer power management scheme for efficient energy delay tradeoff in a WLAN,” in *International J. Comput. Telecommun. Netw.*, vol. 51, pp. 1–6, Oct. 2007.

[12] M. Pereira, L. Bernardo, R. Dinis, R. Oliveira, P. Carvalho, and P. Pinto, “Delay optimization on a p-persistent MAC protocol for a multi-packet detection in SC-FDE system,” in *Proc. 2010 IEEE Wireless Commun. Netw. Conf.*, pp. 1–6.

[13] F. Sivrikaya and B. Yener, “Minimum delay routing for wireless networks with STDMA,” *Wireless Netw.*, vol. 15, no. 6, pp. 755–772, Oct. 2007.

[14] P.-J. Wan, S. C.-H. Huang, L. Wang, Z. Wan, and X. Jia, “Minimumlatency aggregation scheduling in multihop wireless networks,” in *Proc. 2009 MobiHoc*, pp. 185–194.

[15] H. Li, Y. Cheng, C. Zhou, and W. Zhuang, “Minimizing end-to-end delay: a novel routing metric for multi-radio wireless mesh networks,” in *Proc. 2009 IEEE INFOCOM*, pp. 46–54.

[16] A. Alzahrani and M. Woodward, “End-to-end delay in localized qos routing,” in *Proc. 2008 IEEE Singapore International Conf. Commun. Syst.*, pp. 1700–1706.

[17] S. Cui, R. Madan, A. Goldsmith, and S. Lall, “Cross-layer energy and delay optimization in small-scale sensor networks,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 10, pp. 3688–3699, Oct. 2007.

[18] X. Xia, Q. Ren, and Q. Liang, “Cross-layer design for mobile ad hoc networks: energy, throughput and delay-aware approach,” in *Proc. 2006 IEEE Wireless Commun. Netw. Conf.*, vol. 2, pp. 770–775.

[19] A. Pakdehi and F. Ashtiani, “Cross-layer optimization of adaptive modulation and coding preserving packet average delay time,” in *Proc. 2008 IEEE GLOBECOM*, pp. 1–5.

[20] M. Xiao, M. Medard, and T. Aulin, “Cross-layer design of rateless random network codes for delay optimization,” *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3311–3322, Dec. 2011.

[21] C.-C. Wang and N. Shroff, “On wireless network scheduling with intersession network coding,” in *Proc. 2008 Conf. on Inf. Sci. Syst.*, pp. 30–35.

[22] A. Ukil, “Cross-layer optimization in QoS aware next generation wireless networks,” in *Proc. 2009 International Conf. on Inf., Commun. Signal Process.*, pp. 1–5.

[23] A. Vosoughi and Y. Jia, “Maximizing throughput in cooperative networks via cross-layer adaptive designs,” in *Proc. 2010 IEEE Sarnoff Symp.*, pp. 1–6.

[24] M. Cheng, X. Gong, and L. Cai, “Link rate allocation under bandwidth and energy constraints in sensor networks,” in *Proc. 2008 IEEE GLOBECOM*, pp. 1–5.

[25] B. Escoffier, J. Monnot, and V. T. Paschos, “Weighted coloring: further complexity and approximability results,” *Inf. Process. Lett.*, vol. 97, pp. 98–103, 2006.

[26] J. Mycielski, “Sur le coloriage des graphes,” *Colloq. Math.*, pp. 161– 162, 1955.

[27] M. X. Cheng, X. Gong, Y. Xu, and L. Cai, “Link activity scheduling for minimum end-to-end latency in multihop wireless sensor networks,” in *2011 IEEE Globecom*.

II. TRANSMISSION SCHEDULING BASED ON A NEW CONFLICT GRAPH MODEL FOR MULTICAST IN MULTIHOP WIRELESS NETWORKS

Maggie Cheng and Quanmin Ye

Abstract — In multicast applications, the end-to-end delay from the source to a group member is determined by the multicast tree topology and the waiting time at each relay node. This paper addresses when the multicast tree is given how to schedule wireless nodes for transmission so that network delay is minimized. We first model the conflict relation among wireless transmissions in a conflict graph, and then we compute a transmission schedule based on an Integer Linear Programming (ILP) model. Since solving ILP problem is NP-hard, a heuristic is designed to solve the ILP problem. The resulting schedule is conflict-free, which is guaranteed by the feasibility of the ILP model. Simulation results show significant reduction of delay when compared with a First Come First Serve (FCFS) scheduling policy.

I. INTRODUCTION

End-to-end delay is one of the major metrics for quality of service in wireless networks, especially in sensor networks, where sensor nodes are often engaged in a collaborative effort to accomplish a common task. To make sure all destinations receive data timely is important to many sensor network functions, for example, a sink node may send out a query message to a group of sensor nodes, and a mobile sink may frequently update its location with sensor nodes, etc.

In wireless networks with omnidirectional antennas, the signal from one transmitter could reach all neighbors. The broadcast nature of wireless transmissions could be an advantage sometimes in multicast applications where the neighbors are intended receivers, but it could be a disadvantage when they are not. Whether the transmissions from two nodes conflict with each other depends on the location of the receivers— if a receiver of node A is in the interference range of node B, then the transmissions from node A and node B conflict. Transmissions that are conflicting with each other cannot happen simultaneously, therefore they must be scheduled one after another, and this is the main cause of delay. Which one should transmit first is a scheduling issue at the MAC layer. In deterministic scheduling algorithm, the waiting time for a packet at a relay is known from its schedule.

To find the optimal schedule that yields minimum delay is a very challenging problem, more challenging than the unicast version of the problem ([1]). In multicast, routing paths to all destinations form a multicast tree, and the self-interference from the other branches of the same data flow can cause delay. Fig. 1 shows that if there is only one unicast flow from node 1 to node 15, end-to-end delay is 3 slots; However, if there are other destinations, the end-to-end delay from node 1 to node 15 can increase to 5 slots, because the relay nodes on other branches interfere with it. The results shown in Fig. 1 are for minimizing the total delay from node 1 to all destinations (nodes 8– 15). The routing paths (highlighted) are given as the shortest paths. The slot number assigned to a node is labeled besides the node. The transmission schedule is computed using the optimal scheduling algorithm proposed in this paper.

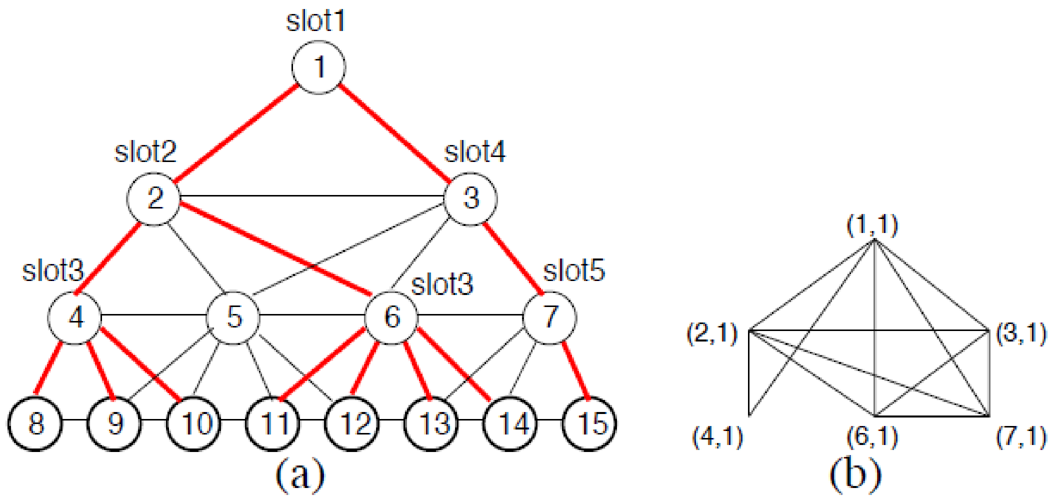


Fig. 1. Multicast tree and conflict graph

In a multicast session, the delay on the shortest path 1–3–7–15 is 5 slots time; whereas in a unicast flow, it is only 3 slots time. (a) the multicast tree, (b) the conflict graph.

When there are multiple data flows in the network, it is more difficult to find the optimal transmission schedule that leads to the minimum end-to-end delay. In this paper, we propose linear programming-based transmission scheduling scheme to compute the time slot assignment for wireless nodes such that the end-to-end delay is minimum. The scope of the paper is on data link layer only, assuming the routing information is given.

The rest of the paper is organized as follows. In Section II, we briefly survey the most related work in reducing delay of multicast and broadcast in wireless networks; in Section III, we present the conflict graph model for multicast, formulate the minimum delay multicast problem as an Integer Linear Program, and propose an algorithm for computing the schedule. In Section IV we compare our algorithms with the commonly used FCFS algorithm; Finally, Section V concludes the paper.

II. RELATED WORK

In many wireless ad hoc and sensor network operations, such as broadcasting a control data packet, or disseminating a query in sensor networks etc., message dissemination from a source node to a group of nodes needs to be as fast as possible, and sometimes it has stringent time requirement. The problem of minimum delay multicast has attracted a lot of research interests. In [2]–[4], the Minimum Broadcast Schedule problem in wireless networks was addressed. It is assumed that nodes all have uniform transmission range, and nodes can use multiple channels to transmit. To avoid collision, a pair of sender and receiver are assigned the same channel for transmission between them while their neighboring nodes use different channels. The proposed work includes to compute a broadcast tree, and to compute an efficient schedule for the transmissions on the tree. Our work uses a different channel model from [2]–[4], where all nodes transmit at the same frequency channel, and therefore a different interference relation is used as the basis for scheduling.

The main interest in minimum delay broadcast has been to compute the broadcast tree, i.e., the routing problem for broadcast. The minimum delay broadcast routing problem in wireless networks was surveyed in [5]. Gandhi et al. ([6]) studied the problem for heterogeneous wireless networks where nodes may have different transmission ranges. Stojmenovic et al. proposed a dominating set based broadcast scheme with neighbor elimination ([7]) for wireless ad hoc networks. Ravi et el. ([8]) studied the minimum delay broadcast problem for the telephone network model, in which a node is only allowed to communicate with at most one of its neighbors at each step. Other work on minimum broadcast time based on the telephone model includes [9]–[11].

In comparison, there are not many publications in minimum delay scheduling for multicast, given the multicast routing information. Chaporkar et al. [12] addressed the MAC layer multicast problem as an instance of the stochastic shortest path problem and developed an optimal transmission strategy for minimum delay multicast. However, [12] is different from our work in the sense that the multicast destinations are all one hop neighbors of the source, and the scheduling issue only concerns with whether the source should transmit to all destinations in one round, or use multiple rounds to finish the task. Our work involves scheduling of nodes in a multihop network, where the destinations of the source are multiple hops away, and the multicast routing is given as input. [13] explores the broadcast scheduling problem in multi-hop wireless networks, in which simultaneous transmissions from neighboring nodes are not allowed. Such a conflict model is static and is not changing with the dynamic traffic of the network. Our work is based on a more accurate dynamic conflict graph model, in which two transmissions are considered conflicting only if the receiver(s) of one transmitter is in the interference range of the other transmitter, so the schedule automatically avoids the hidden terminal problem and exposed terminal problem.

In minimum delay scheduling, the majority in the literature was for scheduling of unicast flows in multihop wireless networks. Through transmission scheduling or transmission probability control, the end-to-end delay can be improved. Chatterjee et al. [14] presented a scheduling scheme for nodes such that the maximum latency to send a packet from a node to the sink is minimized. Sarkar et al. [15] addressed the energy-delay tradeoff problem and formulated the problem as a constrained optimization problem that achieves the minimum energy while satisfying the constraint on average packet delay. The optimization problem was solved by using a dynamic programming formulation, from which a closed-form expression for the optimal sleep duration is derived. Pereira et al. [16] addressed delay optimization problem for a random access MAC protocol. They presented an accurate analytical model to derive the optimal transmission probability of each mobile node that minimizes the delay.

III. TRANSMISSION SCHEDULING

A. Conflict Graph

Given the routing information in a wireless network $G = (V, E)$ we can build its conflict graph G'= (V', E') as follows: each vertex $u \in V'$ is a transmission denoted by an ordered pair (transmitter-id, flow-id), and each edge $(u, v) \in E'$ indicates the conflict relation between vertex u and vertex v. The *transmitter-id* is the identifier of the transmitting node, which could be the source or a relay node. The *flow-id* is the identifier of the source node. A relay node can relay for different sources, so a *flow-id* is necessary to make the distinction.

We assume the link layer ACK is not used upon successful reception of a packet for multicast/broadcast. Two transmissions are considered conflicting with each other if a receiver of one transmitter is in the interference range of the other transmitter. If the conflict graph were built based on the location of transmitters without considering the location of the receivers, it will include edges between all nodes within 2-hop distance. This will result in a dense conflict graph that ultimately wastes channel spectrum. The conflict relation of nodes varies with the multicast trees. Therefore for broadcast/multicast traffic, updating the conflict graph whenever the multicast trees change is necessary.

In the literature, the link-based model is widely used (e.g., [1]), in which each vertex $u \in V'$ represents a wireless link, and each edge in E' indicates the conflict relation between a pair of wireless links. We observed that such a link-based model does not accurately capture the conflict relation in multicast, since each node's transmission correspond to multiple active links, and the conflict relation between a pair of links is not equivalent to the conflict relation between a pair of nodes. Therefore for broadcast/multicast, a node-based conflict model is suitable. The transmission scheduling scheme in the following section is based on such a conflict graph when the routing information is given.

B. Integer Linear Program Model for Scheduling

Let F be the total number of distinct slots in a TDMA frame. Assume each slot time is exactly one packet transmission time, then F is decided by wireless link bandwidth and packet size. F is used as input in the integer linear program. Let variable $d_{v,s,i}$ represent the delay at node v for a packet generated by source s, which includes node v's waiting time before transmission and the mandatory storeand-forward delay. Binary variable $x_{v,s,i}$ is set to '1' when the incoming link uses a slot number larger than the one used by the outgoing link. Binary variable $sl_{v,s,f} = 1$ indicates slot f is assigned to node v for transmitting packets generated by source s. Since multicast routing information is given, we know the multicast tree P_s for source node s, and the routing path $P_{s,d}$ from s to a destination node d. We use constant $Path_{v,s} = 1$ to indicate node v is on the routing paths for source s, and = 0 otherwise. R_s is the source rate, defined as the number of packets generated per frame period.

After we have the conflict graph, the optimization of delay can be modeled as an integer linear program. Let A_s be the packet generation time at source s, we can minimize the total end-to-end delay, including the initial access delay at the source node. The reason we consider initial access delay at the source is because this is for multicast, and therefore the delay at the source will delay the receiving time of multiple destinations, so the initial access delay is too important to ignore. If the objective is to minimize the make-span of the schedule, i.e., the duration from the first node transmits until the last node receives it, we can use the same objective function in the linear program but make $A_s = 0$.

In the following, $v \in P_s$ means v is a transmitting node in the multicast tree of source s; $v \in P_{s,d}$ is a transmitting node on the path from s to d. v could be the source node or a relay node. In (3d), $(u, v) \in P_s$ means edge (u, v) is in the tree, and node u and v both are transmitters. Now we formally present the ILP model as follows.

To minimize the total end-to-end delay of all destinations:

Minimize

$$\sum_{s \in S} \sum_{i=1}^{R_s} \sum_{d \in D_s} \sum_{v \in P_{s,d}} d_{v,s,i} \tag{1}$$

Or to minimize the maximum end-to-end delay:

Minimize

$$\max_{s \in S} \max_{i=1..R_s} (\max_{d \in D_s} (\sum_{v \in P_{s,d}} d_{v,s,i})) \tag{2}$$

Subject to

$$sl_{v,s,f} + sl_{v',s',f'} \leq 1, \forall \big((v,s),(v',s')\big) \in E', \forall f = 1..F \tag{3a}$$

$$\sum_{f=1}^{F} sl_{v,s,f,i} = Path_{v,s}, \forall i = 1..R_s, \forall v \in P_s, \forall s \in S \quad (3b)$$

$$sl_{v,s,f} = \sum_{i=1}^{R_s} sl_{v,s,f,i}, \forall v \in P_s, \forall s \in S, \forall f = 1..F, \quad (3c)$$

$$d_{v,s,i} = \sum_{f=1}^{F} sl_{v,s,f,i}\, f - \sum_{f=1}^{F} sl_{u,s,f,i}\, f + x_{v,s,i}F,$$

$$\forall (u, v) \in P_s, \forall s \in S, \forall i = 1..R_s \quad (3d)$$

$$d_{s,s,i} = \sum_{f=1}^{F} sl_{s,s,f,i}\, f - A_s + x_{s,s,i}F, \forall s \in S, \forall i = 1..R_s \quad (3e)$$

$$0 \le d_{v,s,i} \le F, \forall v \in P_s - \{s\}, \forall s \in S, \forall i = 1..R_s \quad (3f)$$

$$0 \le d_{s,s,i} \le F, \forall s \in S, \forall i = 1..R_s \quad (3g)$$

$$sl_{v,s,f} = \{0,1\}, sl_{v,s,f,i} = \{0,1\}, \forall v \in P_s, \forall s \in S, \forall f = 1..F, \forall i = 1..R_s \quad (3h)$$

$$x_{v,s,i} = \{0,1\}, \forall v \in P_s, \forall s \in S, \forall i = 1..R_s \quad (3i)$$

In the constraint (3a), the conflict graph model is used to make sure the conflicting transmissions cannot be scheduled to use the same slot. (3b -3c) is to make sure the number of slots assigned to a node is proportional to the data rate of the node; (3d -3e) model the delay at each node on a per-packet basis.

1) *LP-Rounding Algorithm for Slot Assignment:* To solve the above integer linear program is NP-hard. We first relax it to a linear program, then use rounding to map real numbers to integers for valid slot assignment.

1) Find the optimal solution for the LP problem with all integer variables relaxed to real numbers;
2) Sort $sl_{v,s,f,i}$ in non-increasing order, set threshold Th = 0.5;

3) Pick the largest $sl_{v,s,f,i}$ from the sorted list. If $sl_{v,s,f,i} > Th$, assign $sl_{v,s,f,i} = 1$ and $sl_{v,s,f} = 1$. Assign $sl_{v's',f}$ and $sl_{v's',f,i}$ for other transmissions (v', s') that are conflicting with the transmission of (v, s). Assign remaining values appropriately to satisfy (3b); If $Th >$ the largest $sl_{v,s,f,i}$, set Th = the largest $sl_{v,s,f,i}$;
4) Repeat step 3) until all variables are rounded to integers.

IV. SIMULATION

The nodes are randomly deployed on a 150m×150m square region. Node transmission range is set to 30m. We only test on the connected networks. Source and destinations are randomly chosen but we also make sure they sit across the network. The multicast tree is given as input. We use the proposed LPbased scheduling scheme to compute the slot assignment, and get the end-to-end delay. The results are compared with the widely used FCFS scheme, in which a node is assigned to use the next available slot as soon as it arrives at a relay node. To obtain a fair comparison, the FCFS scheme also uses the network topology to make sure the new slot assignment has no conflict with existing assignments.

We first observe how the delay changes with increasing source rate. The TDMA frame size is 30 slots, and each slot time is one packet transmission time. If the source generates one packet each frame, then the source rate is 1/30B. We define the baseline rate = 1/30B, where B is the wireless link bandwidth. The plot shows when source rate is 1X and 3X of the baseline rate respectively.

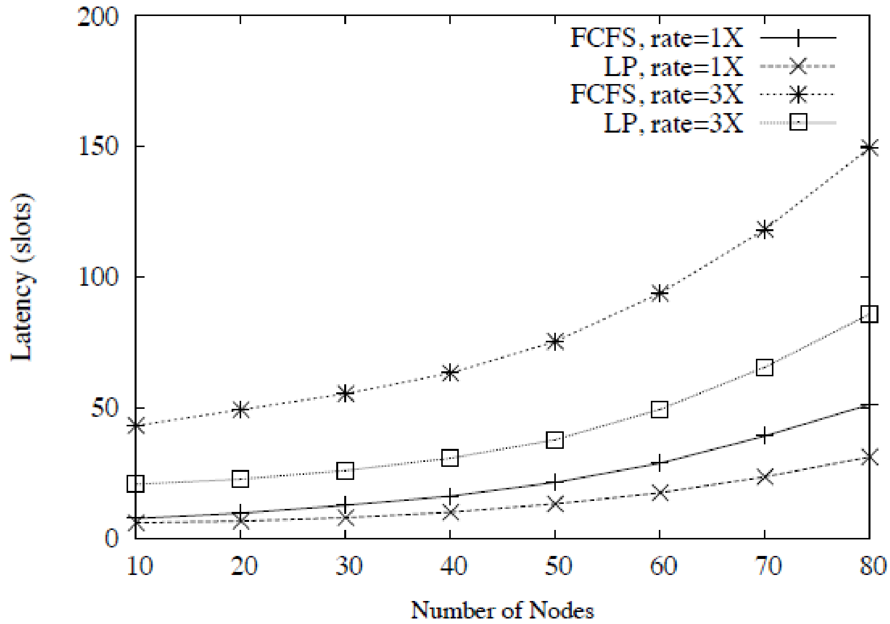


Fig. 2. Comparison of two scheduling schemes under different source rates

The first simulation was done with 2 multicast sessions in the network. Each session has one source node sending to 5 destinations across the network. When the network size increases from 10 to 80 nodes, the average hop counts from the source to its destinations increases from 1.6 to 6.1. Delay increases nonlinearly with hop counts. We observed that the proposed scheduling scheme LP outperforms the FCFS more when the source rate is higher, and when the network size is larger (therefore the path length is larger). In average, the performance gain of LP over FCFS is 22% – 40% for rate=1X, and 43% to 52% for rate=3X. Each data point in the plot shows the average result of 10 randomly deployed networks.

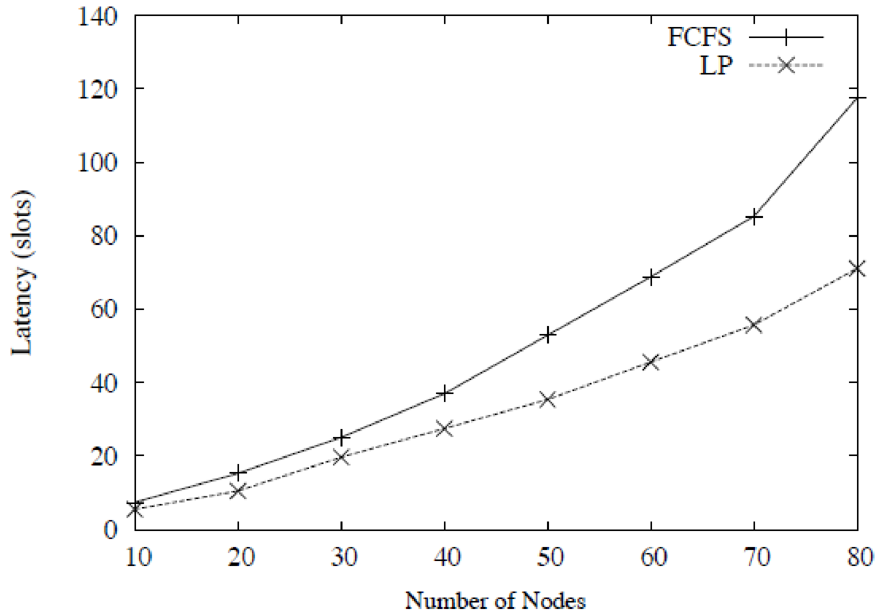


Fig. 3. Comparison of the two scheduling schemes

In the second simulation study, we compare the proposed LP-based scheduling algorithm with FCFS with increasing number of flows. We create a network of n nodes, with |S| sources. Each source has a group of destinations. The average destination group size is 5 nodes for each source. The plot shows delay data for n = 10 & 80, |S| = 0.2n. The source rate is 1X of the baseline rate. We observed that LP-based scheduling scheme outperforms FCFS in every single scenario, and the performance gain is 26% (for 10 nodes) to 40% (for 80 nodes).

V. CONCLUDING REMARKS

In this paper, we study how to achieve the minimum end-to-end delay for multicast applications in a multihop wireless network. In multicast and broadcast, one transmitter may have multiple receivers, and therefore one transmission may correspond to multiple active links. For this reason, the traditional link-based conflict graph model (as in [1]) fails to accurately model the conflict relation in multicast. We propose a node-based conflict graph model, in which two nodes are considered conflicting if and only if at least one of the receivers of one transmitter is in the interference range of the other transmitter. The conflict graph model dynamically changes with traffic, and accurately captures the conflict relation between nodes. Based on the conflict model, we build a linear programming model to compute the schedule of relay nodes. The result shows the proposed scheduling scheme achieves significant delay reduction from the widely used FCFS model.

REFERENCES

[1] M. X. Cheng, X. Gong, Y. Xu, and L. Cai, "Link activity scheduling for minimum end-to-end latency in multihop wireless sensor networks," in *IEEE Globecom'11*, 2011.

[2] W. SHANG, P. WAN, and X. HU, "Approximation algorithms for minimum broadcast schedule problem in wireless sensor networks," *Front. Math. China,* vol. 5, no. 1, pp. 75–87, 2010.

[3] J. Zhu, X. Chen, and X. Hu, "Minimum multicast time problem in wireless sensor networks," in *Wireless Algorithms, Systems, and Applications*, 2006, p. 490501.

[4] J. Zhu, W. Shang, and X. Hu, "New algorithm for minimum multicast time problem in wireless sensor networks," in *Wireless Communications and Networking Conference,* 2007.WCNC 2007. IEEE, march 2007, pp. 3529 –3534.

[5] P. A., *Broadcasting in radio networks*. NewYork:John Wiley and Sons, Inc., 2002, pp. 509 – 528.

[6] R. Gandhi, S. Parthasarathy, and A. Mishra, "Minimizing broadcast latency and redundancy in ad hoc networks," in *Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing*, ser. MobiHoc '03, 2003, pp. 222–232.

[7] I. Stojmenovic, M. Seddigh, and J. Zunic, "Dominating sets and neighbor elimination-based broadcasting algorithms in wireless networks," *IEEE Trans. Parallel Distrib. Syst*., vol. 13, pp. 14–25, January 2002.

[8] R. Ravi, “Rapid rumor ramification: approximating the minimum broadcast time,” in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*. Washington, DC, USA: IEEE Computer Society, 1994, pp. 202–213.

[9] A. Bar-Noy, S. Guha, J. S. Naor, and B. Schieber, “Message multicasting in heterogeneous networks,” *SIAM Journal on Computing*, vol. 30, no. 2, pp. 347–358, 2001.

[10] S. Banerjee, C. Kommareddy, K. Kar, B. Bhattacharjee, and S. Khuller, “Construction of an efficient overlay multicast infrastructure for real-time applications,” in *IEEE INFOCOM 2003*, vol. 2, 2003, pp. 1521 – 1531.

[11] G. Kortsarz and D. Peleg, “Approximation algorithms for minimum-time broadcast,” *SIAM J. Discret. Math*., vol. 8, pp. 401–427, August 1995.

[12] P. Chaporkar and S. Sarkar, “Minimizing delay in losstolerant mac layer multicast,” in Modeling and *Optimization in Mobile, Ad Hoc, and Wireless Networks, 2005. WIOPT 2005. Third International Symposium* on, April 2005, pp. 358 – 367.

[13] E. L. Lloyd, Broadcast scheduling for TDMA in wireless multihop networks. New York, NY, USA: JohnWiley&Sons,Inc.,2002,pp.347–370.[Online].Available: http://dl.acm.org/citation.cfm?id=512321.512337

[14] P. Chatterjee and N. Das, “A cross-layer distributed tdma scheduling for data gathering with minimum latency in wireless sensor networks,” in *Wireless VITAE 2009*, may 2009, pp. 813 – 817.

[15] M. Sarkar and R. L. Cruz, “A mac layer power management scheme for efficient energy delay tradeoff in a wlan,” in *The International Journal of Computer and Telecommunications Networking*, 2007, vol. 51, October 2007, pp. 1 –6.

[16] M. Pereira, L. Bernardo, R. Dinis, R. Oliveira, P. Carvalho, and P. Pinto, “Delay optimization on a p-persistent mac protocol for a multi-packet detection in sc-fde system,” in *Wireless Communications and Networking Conference (WCNC), 2010 IEEE*, April 2010, pp. 1 –6.

III. SIMULTANEOUS ROUTING AND MULTIPLEXING IN AD HOC NETWORKS WITH MIMO LINKS

Maggie Cheng, Quanmin Ye and Xiaochun Cheng

Abstract — this paper addresses how to leverage the spatial multiplexing function of mimo links to improve wireless network throughput. Wireless interference modeling of a half-duplex mimo node is presented, based on which, routing, spatial multiplexing and scheduling are jointly considered in one optimization model. A linear program-based algorithm is proposed for the joint optimization, and numerical simulation results show that the joint optimization of routing with spatial-temporal multiplexing is superior to the separate design approaches, including separating routing from the other two designs, and separating scheduling from the other two designs.

I. INTRODUCTION

Multiple Input Multiple Output (MIMO) technology has brought dramatic change to wireless communication. In the conventional single-input single-output (SISO) transmission mode, a node can only receive from one transmitter and none of the other neighbors should be transmitting while a node is receiving; while in a MIMO mode, a transmitter equipped with multiple antennas can split data into multiple streams and transmit them simultaneously over multiple antennas, and a receiver can also receive data from multiple antennas by suppressing multiuser interference. As long as there is an independent channel between a pair of transmit-receive antennas, the receiver can recover the received data. By using spatial multiplexing a MIMO link can achieve multiple times of the baseline data rate. Spatial multiplexing has been successfully leveraged to improve the capacity of point-to-point communication and infrastructure-based wireless networks such as cellular networks.

It is relatively new to apply MIMO technology in multihop networks, especially in multihop ad hoc networks. In wireless ad hoc networks, every node is a peer, and there is no predetermined routers or relay nodes. The main challenge to use MIMO technology in ad hoc networks is coordinated transmission involving spatial multiplexing. In order to make the best use of the radio capacity, nodes need to cooperatively evaluate the channel condition, and decide how the data packets will be routed and how the data streams will be allocated to transmit-receive antennas. The former is mainly a physical layer issue, while the latter is cross the network layer and lower layers. This paper focuses on the second part of the challenge: how to decide the routing paths and how to allocate data streams to transmit-receive antenna pairs so that network throughput is maximized. It is important the spatial multiplexing function is considered during the routing process so that the radio spectrum is used in an efficient manner. It is also important to consider multiplexing of different data streams in the time domain. An antenna may be used to transmit different data streams at different time. Overall it is a joint design involving routing, antenna allocation and scheduling. A fundamental problem pertaining to the joint design is interference management and interference mitigation.

The rest of the paper is organized as follows. In Section II we review related work in routing and multiplexing in MIMO-based ad hoc networks; in Section III we present the channel model for MIMO links; in Section IV we present interference modeling in MIMObased multihop networks; in Section V we formulate the maximum throughput problem as a mixed integer linear program; in Section V we provide numerical simulation results on the performance gain of our algorithm; Section VI concludes the paper.

II. RELATED WORK

MIMO technology has experienced a recent boom in wireless communications. During the technological boom, the throughput capacity of MIMO systems has become a research topic of rich content, from the capacity of a single-user MIMO channel to multiuser MIMO systems, from single-hop communication to multihop networks.

For a single-user MIMO channel, the water-filling method is known to have achieved the maximum data rate [1]. In a multiuser environment, the link capacity is more difficult to estimate. Studies for multiuser MIMO systems branched out into several subtopics. Previous work [2]–[7] focused the capacity of a multiuser MIMO system in a broadcast channel or single-hop network, such as a cellular network. Liu et al. [8] considered a network consisting of L interfering concurrent transmission pairs using MIMO links, and provides a solution that guarantees a global optimal solution to the maximal sum of mutual information (MSMI) problem in a multiuser MIMO system.

Multiuser diversity has been exploited in cellular networks [9]–[11] to increase the capacity gain. Chu et al. further extended it to multihop ad hoc networks [12]. By exploiting the multiuser diversity and spatial diversity, the scheduling algorithms in [12] opportunistically selects a subset of nodes to serve as transmitters and allocate data streams to antennas based on packet priority, channel quality etc.

Routing as a network layer issue has also been impacted by the use of MIMO links. In [13], MIMOaware routing was proposed that uses a separate-layer approach to address the routing problem, in which the actual capacity of each MIMO channel is estimated on a periodic basis and the statistics is used in routing decision. In [14], channel-aware MAC protocol design was applied to ad hoc networks, in which spatial diversity is explored to combat fading and achieve robustness. [15] used a SINR model and then developed a CSMA-based MIMO-pipe scheduling under the SINR model. Physical layer provides the upper layers a set of rates and SINR requirements, which capture the rate-reliability tradeoff in MIMO communications. Under the physical model, transmission on a link is said to be successful if its SINR is greater than a pre-determined threshold. Compared to the widely used protocol model the physical model can more accurately capture the channel condition and the intrinsic probabilistic nature of wireless communications [16]. [17], [18] also used physical interference models for scheduling.

III. MIMO CHANNEL MODEL

A MIMO link can operate at the spatial multiplexing mode to increase its data rate or at the diversity coding mode to improve link diversity or reliability. MIMO technology provides a node with the ability to suppress multiuser interference (MUI), so a node can receive from multiple transmitters at the same time.

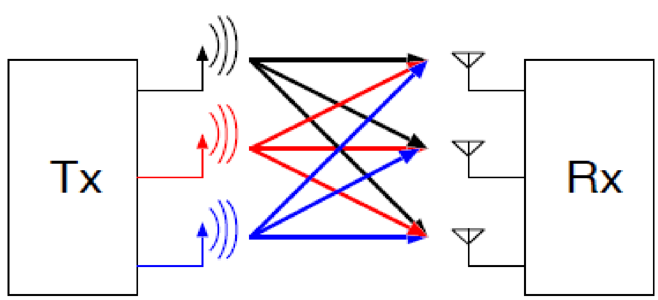


Fig. 1. MIMO Channel Model

Fig. 1 shows the MIMO channel model. The mathematical description of a narrow-band flat-fading MIMO channel with multiple transmit and receive antennas is ([19], 20]):

$$y = Hx + n$$

Where y is the receive signal vector, x is transmit signal vector, n is noise vector, and H is the channel matrix. H_{ij} is the spatial channel coefficient between the ith antenna of the receive node and the jth antenna of the transmit node, and it is the sum of the line of sight (LOS) component and the fading component. If the environment is scattering enough, and receive antennas are sufficiently apart, there are $min\{N_t, N_r\}$ independent transmission paths, where N_t and N_r are the number of transmit antennas and receive antennas respectively. The multiple transmit signals can be the same data bit or different data bits. If different data bits are transmitted from multiple antennas, the total data rate will be multiple times of the baseline data rate. This is called spatial multiplexing.

IV. INTERFERENCE MODELING OF MIMO LINK

The interference relation among MIMO links is fundamentally different from SISO links. With single-radio single-channel links, we can decide whether two links can be active at the same time just by looking at the conflict graph; but with MIMO links, whether two links can be active at the same time cannot be decided without knowing antenna assignment. A routing scheme must jointly consider antenna assignment in order to achieve the maximum network throughput. To allow joint design of routing and spatial multiplexing, the multiplexing capacity of MIMO links must be considered.

We assume a half-duplex channel is used, i.e., a node is either transmitting or receiving, or being idle (e.g., [12]) at any time instance. Let boolean variable $T_{ij} = 1$

indicate node i is transmitting at antenna j; Let $R_{ij} = 1$ indicate node i is receiving at antenna j. A node is transmitting as long as one of its antennas is transmitting; and is receiving as long as one of its antennas is receiving. Let $T_i = 1$ indicate node i is transmitting; Let $R_i = 1$ indicate node i is receiving. A half-duplex MIMO channel can be described as:

$$T_i \geq T_{ij}\ \forall j\ , \forall i \qquad (1a)$$

$$R_i \geq R_{ij}\ \forall j\ , \forall i \qquad (1b)$$

$$T_i + R_i\ \leq 1, \forall i \qquad (1c)$$

In addition to being half-duplex, a node also has the degree of freedom (DOF) constraint. Among all received streams, some are targeted at node i, some are not, and thus considered as interference streams to node i. A node can successfully decode the received data streams if there are enough information about the channel state information between the receiving antennas and transmitting antennas, and the total received streams (targeted for it or not) is bounded by the number of receive antennas. Let N_i denote node i's one-hop neighbors. For node i with Ai antennas, at any time instant, if $R_i = 1$, it is required that

$$\sum_{j \in N_i} \sum_k T_{jk} \quad \leq A_i \qquad (2)$$

Inequality (2) indicates the total data streams targeted at node i and interference streams to node i is bounded by the number of receiving antennas A_i. This is called degree of freedom (DOF) [21]. When overloading is possible with the channel condition, it is bounded above by $A_i(1 + \alpha)$, where α is the overload factor [22].

If node i is transmitting, it is required that

$$\sum_j T_{ij} \leq A_i \qquad (3)$$

V. JOINT DESIGN OF ROUTING AND SPATIAL-TEMPORAL MULTIPLEXING

Routing is concerned with computing the path a flow will take, spatial multiplexing is concerned with assigning flows to antennas, and scheduling is concerned with assigning time slots to transmitting antennas, a.k.a. temporal multiplexing. In previous work [23], a spatial multiplexing and scheduling scheme is proposed when routing is given. We observed that limiting the flow to a predetermined path may not be optimal for the network throughput. A larger optimization space exists that allows joint optimization of routing, spatial multiplexing and scheduling.

For joint design of routing, spatial multiplexing and scheduling, we will decide for each flow which link will be on the routing path, which antenna of a node will transmit the flow, and which time slot it will use by solving one optimization problem. To add routing in the scope, we introduce a new variable $r_{ij,s}$. $r_{ij,s}$> 0 indicate link (i, j) is used to transmit flow s at a rate of $r_{ij,s}$. Let variable $r_{s,s}$be the source rate we try to maximize. Set $r_{i,s} = r_{s,s}$ if i is the source, and set $r_{i,s} = -r_{s,s}$ if i is the destination, and =0 if i is neither the source nor the destination. For the rest, we use a flow network model to solve it. When Routing is not given, the solutions to $r_{ij,s}$ in the following linear program indicate which link will be used for flow from source s.

From the interference modeling in section IV, we add a time dimension to the variables T_{ij} and R_{ij} : $T_{ij,t}$ = 1 if node i is transmitting at antenna j at slot t, and $R_{ij,t}$ = 1 if node i is receiving at antenna j at slot t. We add another dimension to indicate the source of the flow: $T_{ij,t,s}$ = 1 (or $R_{ij,t,s}$= 1) indicates node i is transmitting (or receiving) flow s at antenna j at slot t.

Constant A_i is the number of antennas node i has. With the objective of maximizing the total data rates from all sources $\sum_s r_s$, we can formulate the following linear program:

Maximize

$$\sum_{s} r_{s,s} \tag{4}$$

Subject to

$$\sum_{j\in N_i} \left(r_{ij,s} - r_{ji,s}\right) = r_{i,s}, \forall i, \forall s \tag{5a}$$

$$\sum_{t}\sum_{k} T_{ik,t,s} = \sum_{j} r_{ij,s}, \forall i, \forall s \tag{5b}$$

$$\sum_{k} T_{ik,t,s} = \sum_{k} R_{jk,t,s}, \forall s, \forall link(i,j) \in Path_s, \forall t \tag{5c}$$

$$\sum_{s} T_{ij,t,s} = T_{ij,t}\ \forall i, \forall j, \forall t \tag{5d}$$

$$\sum_{s} R_{ij,t,s} = R_{ij,t}\ \forall i, \forall j, \forall t \tag{5e}$$

$$T_{i,t} \geq T_{ij,t}, \forall j, \forall i, \forall t \tag{5f}$$

$$R_{i,t} \geq R_{ij,t}, \forall j, \forall i, \forall t \tag{5g}$$

$$T_{i,t} + R_{i,t} \leq 1, \forall i, \forall t \tag{5h}$$

$$\sum_{j} T_{ij,t} \leq A_{i,} \forall i, \forall t \tag{5i}$$

$$\sum_{j\in N_i}\sum_{k} T_{jk,t} \leq A_i + \left(1 - R_{i,t}\right) N_\infty\ \forall i, \forall t \tag{5j}$$

$$r_s, r_{ij,s}, r_{ij} \geq 0, T_{i,t}, R_{i,t}, T_{ij,t}, R_{ij,t}, T_{ij,t,s}, R_{ij,t,s} \in \{0,1\} \tag{5k}$$

The objective function is to maximize the total network throughput. Constraint (5a) is for flow conservation. (5b) requires that for each flow the total data rate of node i from all transmit antennas from all time slots should be the same as the total data rate transmitted by node i. (5c) indicates at any time if node i is transmitting flow s, the number of transmit antennas for flow s is the same as the number of receive antennas from i's neighbors. For instance, if node i uses 3 antennas to transmit flow s, there must be 3 antennas receiving flow s at node i's neighbors. (5d) and (5e) indicate a transmit antenna or a receive antenna is active as long as it is transmitting or receiving any flow. (5f)–(5h) are the requirement for being a half-duplex model at slot level. (5i) and (5j) are for the degree of freedom constraints. N_∞ in (5j) is a very large positive number. When $R_{i,t}$ = 1, node i is receiving at time t, the degree of freedom constraint specified in (2) is enforced at time t; when $R_{i,t}$= 0, then constraint (19j) is automatically satisfied.

VI. SIMULATION

A. Cross Layer Design Performance Evaluation

In this paper we proposed a cross-layer design solution by jointly computing routing path and spatialtemporal multiplexing. It is joint design of physical layer (spatial multiplexing), link layer (scheduling) and network layer (routing). We show the performance gain over the previous work in [23], in which Shortest Path Routing is used. We assume there is large enough antenna separation with low correlation between channels. For node distribution, we assume nodes are randomly positioned in a square region of 150m × 150m. Each node can communicate with other nodes in 30m range. Network size ranges from 10 to 80 nodes, among which 20% nodes are sources. Source-destination pairs are randomly chosen.

We assume the baseline data rate is 30 units per second. We compare the throughput of different schemes measured as multiples of baseline data rate. The joint design shows performance gain in all scenarios, with more significant improvement in larger network sizes. This is because when the network size increases, there are enough choices for different routing algorithms to find different routing paths. Fig. 2 shows the performance comparison between the joint design given by the linear program (4)-(5k) over the previous work in [23]. Fig. 2(a) shows results when nodes are equipped with 2 transmit antenna and 2 receive antenna, (b) and (c) are for 3 antennas and 4 antennas respectively. It is easy to see that as the number of antennas on each node increases, the network throughput increases.

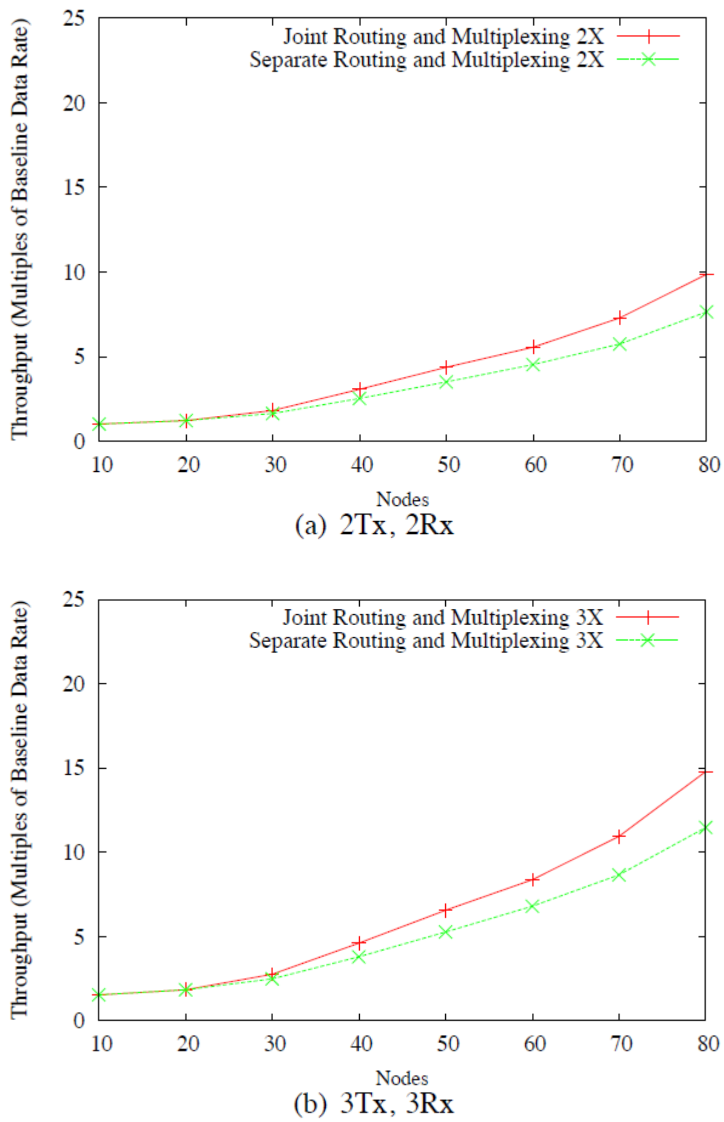
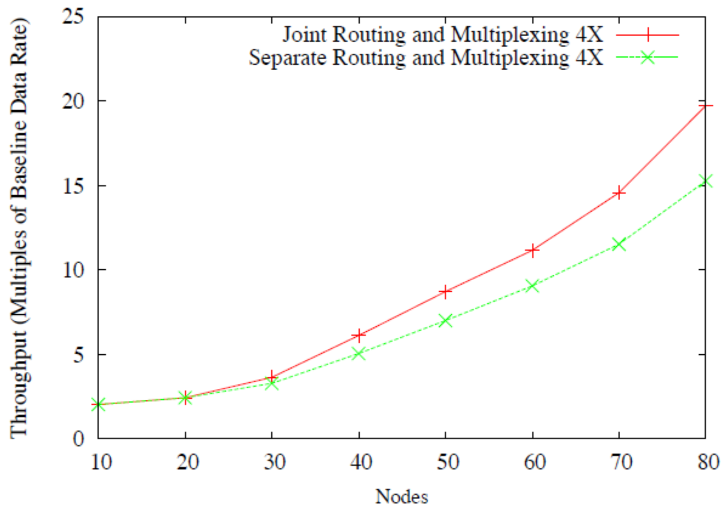


Fig. 2. Joint design vs. fixed shortest path



(c) 4Tx, 4Rx

Fig. 2. Joint design vs. fixed shortest path (cont.)

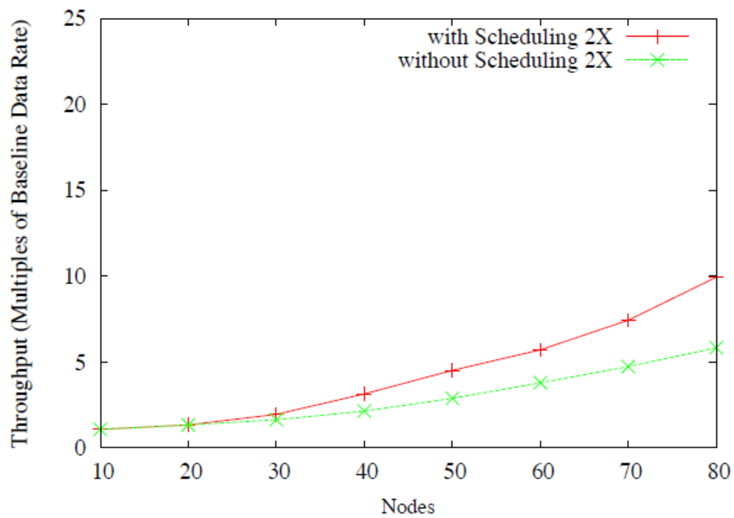
Joint design of routing and spatial-temporal multiplexing vs.fixed shortest path routing with spatial-temporal multiplexing

B. Routing and Spatial-Multiplexing without Scheduling

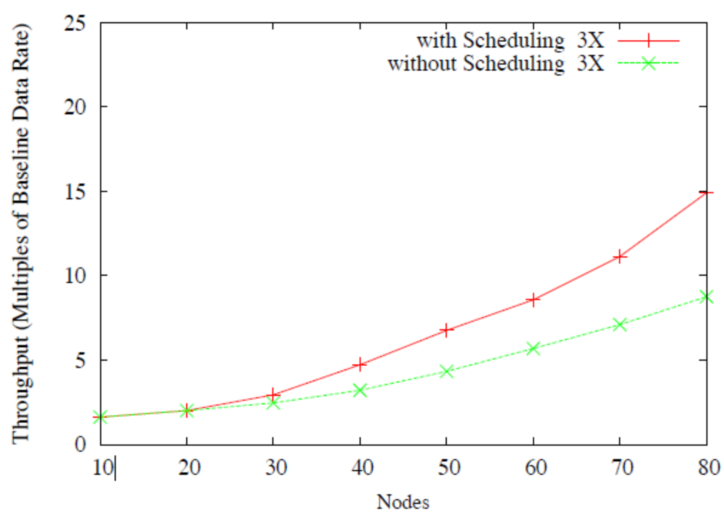
Adding time slot assignment to the scope of crosslayer design certainly can make better use of the channel capacity, but it comes at a cost of higher computational complexity. In this simulation we observe how much performance gain is lost when removing scheduling from the joint design. We compare the throughput obtained by routing+spatial multiplexing with scheduling and the throughput obtained by routing+spatial multiplexing without scheduling.

Without detailed time slot assignment, we assume each link will use all available antenna for transmission and stop transmission when its allocated time is over. The total channel occupancy time is allocated to links to make sure there is no conflict between active transmissions.

The linear program for routing and spatialmultiplexing without scheduling is omitted here. The simulation results are presented in Fig. 3. It shows that when the network size is small, the difference between the two schemes is small; when the network size increases, the throughput gain increases faster with scheduling. Joint design with scheduling increases network throughput by 42% when network has 80 nodes.

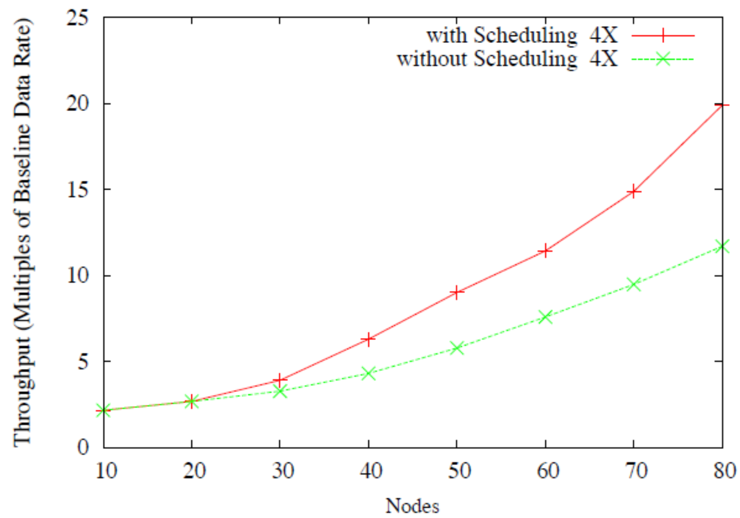


(a) 2Tx, 2Rx



(b) 3Tx, 3Rx

Fig. 3. R+S with Scheduling vs. R+S without Scheduling



(c) 4Tx, 4Rx

Fig. 3. R+S with Scheduling vs. R+S without Scheduling (cont.)

Routing and spatial multiplexing with scheduling vs. routing and spatial multiplexing without scheduling.

VII. CONCLUDING REMARKS

While routing and scheduling have been extensively studied in wireless ad hoc networks, the scope of study has dramatically changed in a MIMO-based ad hoc networks. To add a spatial multiplexing component in the scope creates new challenges and new opportunities. To see the full advantage of MIMO technology, we argue that the spatial multiplexing function should be considered during routing. We propose to jointly consider routing, spatial multiplexing and scheduling at the same time. A linear programming based algorithm is proposed that includes three design problems in one optimization framework. Simulation results verified the advantage of using joint optimization over separate design approaches.

REFERENCES

[1] R. G. Gallager, *Information Theory and Reliable Communication*.Wiley, New York, NY, 1968.

[2] S. Toumpis and A. J. Goldsmith, “Capacity regions for wireless ad hoc networks,” *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*, vol. 2, no. 4, pp. 736–748, July 2003.

[3] R. S. Blum, “Mimo capacity with interference,” *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 793–801, June 2003.

[4] S. Ye and R. S. Blum, “Optimized signaling for mimo interference systems with feedback,” *IEEE Trans. Signal Processing*, vol. 51, no. 11, pp. 2839–2848, Nov. 2003.

[5] S. Cartreux, L. J. Greenstein, and P. F. Dressen, “Simulation results for an interference-limited multiple-input multiple-output cellular system,” *IEEE Commun. Lett.*, vol. 4, no. 11, pp. 334–336, Nov. 2000.

[6] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, “Iterative waterfilling for gaussian vector multiple-access channels,” *IEEE Trans. Inform. Theory*, vol. 50, no. 1, pp. 145–152, Jan. 2004.

[7] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, “Sum power iterative water-filling for multi-antenna gaussian broadcast channels,” *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2004.

[8] J. Liu, Y. T. Hou, Y. Shi, and H. D. Sherali, “Optimization of multiuser mimo networks with interference,” in *GLOBECOM’ 2006*, 2006.

[9] W. Ajib and D. Haccoun, “An overview of scheduling algorithms in mimo-based fourth-generation wireless systems,” *IEEE Netw.*, vol. 19, no. 5, pp. 43–48, Sep.–Oct. 2005.

[10] V. K. N. Lau, Y. Liu, and T. A. Chen, “Optimal multi-user space time scheduling for wireless communications,” in *IEEE VTC 2002-Fall, 2002*, pp. 1939–1942.

[11] D. Aktas and H. E. Gamal, “Multiuser scheduling for mimo wireless systems,” in *IEEE VTC 2003-Fall*, 2003, pp. 1743–1747.

[12] S. Chu and X. Wang, “Opportunistic and cooperative spatial multiplexing in mimo ad hoc networks,” *IEEE/ACM TRANSACTIONS ON NETWORKING*, vol. 18, no. 5, pp. 1610–1623, 2010.

[13] ——, “Mimo-aware routing in wireless mesh networks,” in *IEEE INFOCOM 2010*, 2010.

[14] M. Hu and J. Zhang, “Mimo ad hoc networks: Medium access control, saturation throughput, and optimal hop distance,” *J. Commun. Netw., Special Issue on Mobile Ad Hoc Networks*, vol. 6, no. 4, pp. 317–330, 2004.

[15] D. Qian, D. Zheng, J. Zhang, and S. N, “Csma-based distributed scheduling in multi-hop mimo networks under sinr model,” in *INFOCOM*, 2010.

[16] G. Brar, D. Blough, and P. Santi, “Computationally efficient scheduling with the physical interference model for throughput improvement in wireless mesh networks,” in *in MobiCom, ACM*, 2006, pp. 2–13.

[17] J. Liu, Y. T. Hou, Y. Shi, H. D. Sherali, and S. Kompella, “On the capacity of multiuser mimo networks with interference,” *IEEE Transactions on Wireless Communications*, vol. 7, pp. 488–494, 2008.

[18] B. Hamdaoui and K. G. Shin, “Characterization and analysis of multihop wireless mimo network throughput,” in *MobiHoc*, 2007.

[19] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, and M. Andrews, “An overview of limited feedback in wireless communication systems,” *IEEE Journal on Selected Areas Communications*, vol. 26, pp. 1341–1365, 2008.

[20] A. Tulino, A. Lozano, and S. Verd, “Impact of antenna correlation on the capacity of multiantenna channels,” *IEEE Transactions on Information Theory*, vol. 51, pp. 2491–2509, 2005.

[21] D. Tse and P.Viswanath, *Fundamentals ofWireless Communication. Cambridge*. U.K.: Cambridge Univ. Press, 2005.

[22] B. Vucetic and J. Yuan, *Space-Time Coding*. New York: Wiley, 2003.

[23] M. Cheng and Q. Ye, “A combinatorial solution for scheduling spatial multiplexing in mimo-based ad hoc networks,” in *IEEE Globecom 2012, MIMOCR workshop*, 2012.

IV. A COMBINATORIAL SOLUTION FOR SCHEDULING SPATIAL MULTIPLEXING IN MIMO-BASE AD HOC NETWORKS

Maggie Cheng and Quanmin Ye

Abstract — in a mimo-based ad hoc network, the conflict relation between transmissions is fundamentally changed due to the multiple packet reception capability of mimo nodes. A mimo node can receive from multiple data streams at the same time, which significantly increases data throughput. To fully utilize the additional spectrum capacity enabled by the mimo technology, it is important that transmissions from different nodes are coordinated. The coordination scheme needs to consider both data stream multiplexing and transmission time scheduling for a maximum performance gain. In this paper, we use an optimization framework to jointly consider mimo link spatial multiplexing and scheduling while the routing information is given. A linear program-based algorithm is proposed, and simulation results show it is advantageous over the spatial multiplexing scheme without joint design of scheduling.

I. INTRODUCTION

Multiple Input Multiple Output (MIMO) technology has experienced a recent boom in wireless communications. Different from the single-input single-output (SISO) transmission mode, a node equipped with multiple antennas can split data into multiple streams and transmit them simultaneously over multiple antennas. As long as there is an independent channel between a pair of transmitreceive antennas, the receiver can recover the received data. By using multiple transmit antennas at the transmitter and multiple receive antennas at the receiver, a MIMO link can achieve multiple times of the baseline data rate. This is the spatial multiplexing function of MIMO links.

In the past few years, although MIMO technology has been successfully used for point-to-point communication and infrastructure-based wireless networks (i.e., cellular networks), it has not been widely used in multiple-hop networks such as wireless ad hoc networks. In wireless ad hoc networks, every node is a peer, and there is no predetermined router or relay node. The main challenge to use MIMO technology in ad hoc networks is coordinated transmission involving spatial multiplexing. In order to make the best use of the radio capacity, nodes need to cooperatively evaluate the channel condition and decide how the data streams will be allocated to transmit-receive antennas. This paper focuses on the second part of the challenge: how to allocate data streams to antennas so that the network throughput is maximized.

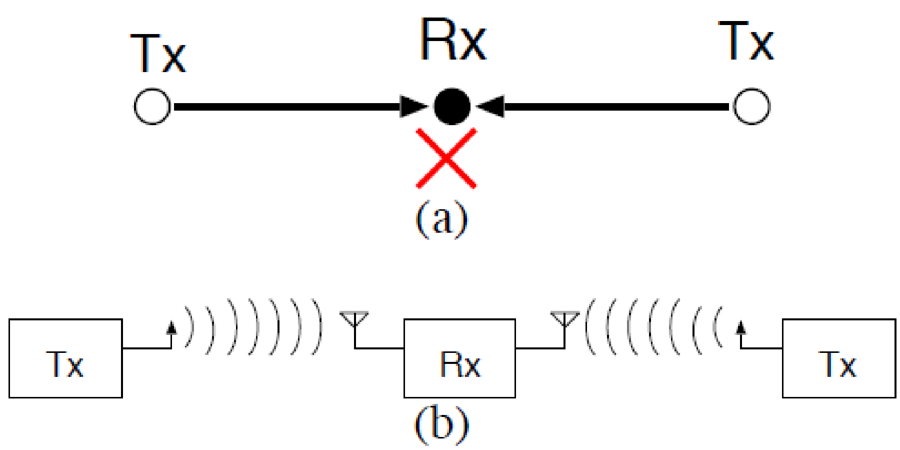


Fig. 1. Hidden Terminal Problem and MIMO Technology

(a) Hidden Terminal Problem; (b) a MIMO receiver can receive two streams

When transforming from a single-radio single-channel ad hoc network to a MIMO-based ad hoc network, a single channel is replaced by multiple spatially multiplexed channels between two nodes, so the conflict relation among transmissions is fundamentally changed. For example, what is considered as a hidden terminal problem (see Fig. 1(a)) in a single-radio single-channel environment is no longer considered a problem with MIMO nodes. In Fig. 1(b), the receive node has two antennas, and there is an independent transmission path between each transmit-receive pair, so the receiver can recover two streams of data simultaneously. Fig. 1 shows the fundamental reason that the MIMO links can increase the network capacity. However, that does not mean there is no limit on the number of transmissions that can be allowed simultaneously. There are new constraints specific to MIMO links, as shown in the section III. How to efficiently make use of the radio spectrum with multiple packet reception capability is a new paradigm to explore. To this end, interference management is a fundamental problem. We will address the interference management problem in spatial multiplexing and transmission scheduling in a joint optimization model, in which interference mitigation is implicitly accomplished through constrained optimization.

The rest of the paper is organized as follows. In Section II, we briefly survey the most related work in MIMO-based ad hoc networks; in Section III, we present interference modeling in MIMO-based multihop networks, and formulate the maximum throughput problem as an Integer Linear Program. In Section IV we validate our model and compare our algorithm with the multiplexing scheme without joint design of scheduling; Finally, Section V concludes the paper.

II. RELATED WORK

MIMO systems have been given extensive study on throughput maximization. For the capacity of the single-user MIMO channel, it is well known that the water-filling method can achieve the maximum data rate [1]. Previous work [2]–[5] have addressed the capacity of multiuser MIMO in a broadcast channel or single-hop network, such as a cellular network. Liu et al. [6] considered a network consisting of L interfering concurrent MIMO transmission pairs (links), and provided a solution that guarantees a global optimal solution to the maximal sum of mutual information (MSMI) problem in a multiuser MIMO system. However none of the aforementioned work directly addresses multihop throughput in a MIMO ad hoc network.

In an effort to improve the network throughput, multiuser diversity has been exploited. [7]–[9] addressed multiuser diversity in cellular networks, and Chu et al. [10] further extend it to multihop ad hoc networks. By exploiting the multiuser diversity and spatial diversity, the scheduling algorithms in [10] opportunistically selects a subset of nodes to serve as transmitters and allocate data streams to antennas based on packet priority, channel quality etc.

In a multiuser environment, the link capacity is more difficult to understand than in a singleradio single-channel system. In [11], MIMO-aware routing was proposed that uses a separate-layer approach to address the routing problem, in which the actual capacity of each MIMO channel is estimated on a periodic basis and the statistics is used in routing decision. In [12], channel-aware MAC protocol design was applied to ad hoc networks, in which spatial diversity is explored to combat fading and achieve robustness. Qian et al. [13] used a SINR model and then developed a CSMA-based MIMO-pipe scheduling under the SINR model. Physical layer provides the upper layers a set of rates and SINR requirements, which capture the rate-reliability tradeoff in MIMO communications. Under the physical model, transmission on a link is said to be successful if its SINR is greater than a pre-determined threshold. Compared to the widely used protocol model the physical model can more accurately capture the channel condition and the intrinsic probabilistic nature of wireless communications [14]. [15] also used physical interference models for scheduling.

In additional to channel-aware routing and MAC protocol design, some previous work considered joint design of the MAC layer and the PHY layer. In [16], a cross-layer design for routing, power allocation and bandwidth allocation was proposed, but the problem is decoupled as a network layer subproblem (for routing) and a link layer subproblem (power allocation to multiple antennas of a node and bandwidth allocation) and is not solved as one optimization problem. In [17], the physical layer (spatial multiplexing) and Medium Access Control strategies are integrated to maximize the network throughput. In [18], the interaction between MAC layer and PHY layer was analyzed and an example cross-layer scheme was presented, however, no concrete optimization framework was given. For example, a node in the network can decide how many antennas to use for transmission depending on the intended receivers distance. A transmit node can use more antennas for higher data rate (or reliability) but shorter distance, or for longer distance and lower data rate, since the node is power constrained. In [17] packets are scheduled at per user level without mixing different users.

III. SPATIAL MULTIPLEXING

A MIMO link uses multiple antennas at both the transmitter and the receiver. It can operate at the spatial multiplexing mode in order to achieve high spectrum efficiency, or in the diversity coding mode in order to achieve better link diversity or reliability. Due to its capability to suppress multiuser interference (MUI), what is considered as conflicting transmissions with SISO links may not be considered conflicting with MIMO links. In Fig.1, the receiver node in the middle can receive two data streams at the same time and still be able to recover both data streams.

A. Interference Modeling of MIMO Links

The interference relation among MIMO links is fundamentally different from that of SISO links. With single-radio single-channel links, we can decide whether two links can be active at the same time just by looking at the conflict graph; but with MIMO links, whether two links can be active at the same time cannot be decided without knowing antenna assignment. This motivates a combinatorial solution for spatial multiplexing in order to achieve the maximum network throughput.

Although transmitting and receiving at the same time is possible by using sophisticated interference suppression scheme, a half-duplex channel is the most commonly used model, in which a node is either transmitting or receiving, or being idle (e.g., [10]). Let $\mathrm{T}_{ij} = 1$ indicate node i is transmitting at antenna j; Let $R_{ij} = 1$ indicate node i is receiving at antenna j. A node is transmitting as long as one of its antennas is transmitting; and is receiving as long as one of its antennas is receiving. Let $\mathrm{T}_i = 1$ indicate node i is transmitting; Let $\mathrm{R}_i = 1$ indicate node i is receiving. T_i, R_i, T_{ij} , R_{ij} are all Boolean variables. A half-duplex MIMO channel can be described as:

$$T_i \geq T_{ij} \ \forall j\ ,\forall i \qquad (1a)$$

$$R_i \geq R_{ij} \ \forall j\ ,\forall i \qquad (1b)$$

$$T_i + R_i \ \leq 1, \forall i \qquad (1c)$$

In addition, a node also has the degree of freedom (DOF) constraint: among all received streams, some are targeted at node i, some are not, and thus considered as interference streams to node i. A node can successfully decode the received data streams if there are enough information about the channel state information between the receiving antennas and transmitting antennas, and the total received streams (targeted for it or not) is bounded by the number of receive antennas. Let N_i denote node i's one-hop neighbors. For node i with $\mathrm{A_i}$ antennas, at any time instant, if $R_i = 1$, it is required that

$$\sum_{j \in N_i} \sum_{k} T_{jk} \quad \leq A_i \qquad (2)$$

Inequality (2) indicates the total data streams targeted at node i and interference streams to node i is bounded by the number of receiving antennas A_i. This is called degree of freedom (DOF) [19]. When overloading is possible with the channel condition, it is bounded above by $A_i(1 + \alpha)$, where α is the overload factor [20].

If node i is transmitting, it is required that

$$\sum_{j} T_{ij} \leq A_i \qquad (3)$$

B. Joint Design of Spatial Multiplexing and Transmission Scheduling

Spatial multiplexing is to assign flows to antennas, and scheduling is to assign time slots to transmitting antennas. Obviously if we do not consider slot assignment while computing spatial multiplexing solution, we can only get a coarsegrained estimation of the achievable throughput. To consider spatial multiplexing at the slot level gives us more performance gain than considering it at the flow level. This motivates the joint design of spatial multiplexing and scheduling.

For joint design of spatial multiplexing and scheduling, we will decide which antenna of a node will transmit a flow and which time slot it will use by solving one optimization problem. Let rs denote the source data rate. We set $r_{ij,s} = r_s, \forall (i,j) \in Path_s$ and $r_{ij,s} = 0$, if (i,j) is not on the routing path. So $r_{ij,s}$ and r_s will be used as scalar variables only. From the interference modeling in section III-A, we add a time dimension to the variables T_{ij} and R_{ij} , $\mathrm{T_i}$ and R_i: $T_{ij,t} = 1$ if node i is transmitting at antenna j at slot t, and $R_{ij,t} = 1$ if node i is receiving at antenna j at slot t; $T_{i,t} = 1$ if node i is transmitting at slot t, $R_{i,t} = 1$ if node i is receiving at slot t. Constant A_i is the number of antennas node i has. With the objective of maximizing the total data rates from all sources we can formulate the following linear program:

Maximize

$$\sum_{s} r_s \tag{4}$$

Subject to

$$r_{ij,s} = r_s, \forall s, \forall (i,j) \in Path_s \tag{5a}$$

$$\sum_t \sum_k T_{ik,t,s} = \sum_j r_{ij,s}, \forall i, \forall s \tag{5b}$$

$$\sum_k T_{ik,t,s} = \sum_k R_{jk,t,s}, \forall s, \forall (i,j) \in Path_s, \forall t \tag{5c}$$

$$\sum_s T_{ij,t,s} = T_{ij,t} \, \forall i, \forall j, \forall t \tag{5d}$$

$$\sum_s R_{ij,t,s} = R_{ij,t} \, \forall i, \forall j, \forall t \tag{5e}$$

$$T_{i,t} \geq T_{ij,t}, \forall j, \forall i, \forall t \tag{5f}$$

$$R_{i,t} \geq R_{ij,t}, \forall j, \forall i, \forall t \tag{5g}$$

$$T_{i,t} + R_{i,t} \leq 1, \forall i, \forall t \tag{5h}$$

$$\sum_j T_{ij,t} \leq A_{i,} \forall i, \forall t \tag{5i}$$

$$\sum_{j \in N_i} \sum_k T_{jk,t} \leq A_i + \left(1 - R_{i,t}\right) N_\infty \, \forall i, \forall t \tag{5j}$$

$$r_s, r_{ij,s}, r_{ij} \geq 0, T_{i,t}, R_{i,t}, T_{ij,t}, R_{ij,t}, T_{ij,t,s}, R_{ij,t,s} \in \{0,1\} \tag{5k}$$

Inequality (5j) is the degree of freedom constraint equivalent to (2). A large positive number N_∞ is used in order to make sure that when node i is not receiving at time t (i.e., $R_{i,t} = 0$) the inequality is automatically satisfied, and therefore the degree of freedom constraint is not enforced on a non-receiving node.

IV. SIMULATION

The objectives of the simulation are to validate the optimization model, and to show the advantage of joint design of spatial-temporal multiplexing over spatial multiplexing alone.

Without a detailed schedule for antennas, we assume each link will use all available antennas for transmission and stop transmission when its allocated time is over. The total channel occupancy time is allocated to links to make sure there is no conflict between active transmissions. The integer linear program can be formulated as follows:

Maximize

$$\sum_{s} r_s \tag{6}$$

Subject to

$$r_{ij,s} = r_s, \forall(i,j) \in Path_s, \forall s \tag{7a}$$

$$r_{ij} = \sum_{s} r_{ij,s} \ , \forall(i,j) \tag{7b}$$

$$\frac{r_l}{A_l B} + \sum_{c=1}^{C} \max_{k:k\in V_c', k\in N_l'} \left\{\frac{r_k}{A_k B}\right\} \leq 1, \forall l \in V' \tag{7c}$$

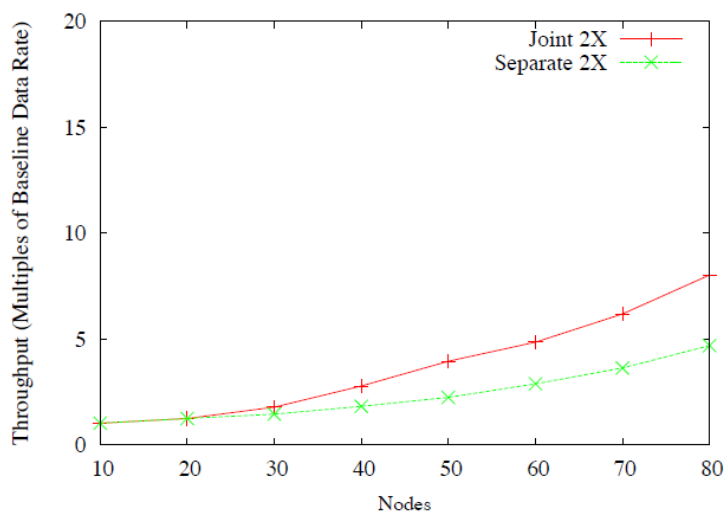
$$r_s, r_{ij,s}, r_{ij} \geq 0, \forall s, \forall(i,j) \ \in E \tag{7d}$$

Inequality (7c) is the key constraint to make sure there is no conflict among transmissions. $\frac{r_l}{A_l B}$ is the fraction of channel occupancy time. Note the conflict graph is built for directional edges, inequality (7c) is applied to each vertex l in the conflict graph $G' = (V', E')$. In constraint (7c), l is a wireless link from node i to node j, and $A_l = \min\{A_i, A_j\}$, where A_i and A_j, are the numbers of antennas of node i and node j, respectively.

To validate the optimization model we use two straightforward cases. The network consists of three nodes, A, B and C, as shown in Fig. 2. In case (1), node A sends to node C via node B; in case (2), there are two data streams—A sends to C via B and C sends to A via B. The baseline data rate is 30 packets per frame time (B=30). We assume each node has three antennas. In both of cases, the joint design of spatial-temporal multiplexing and the separate design both agree with the optimal solution, and achieve a total throughput of 45 packets per frame time.

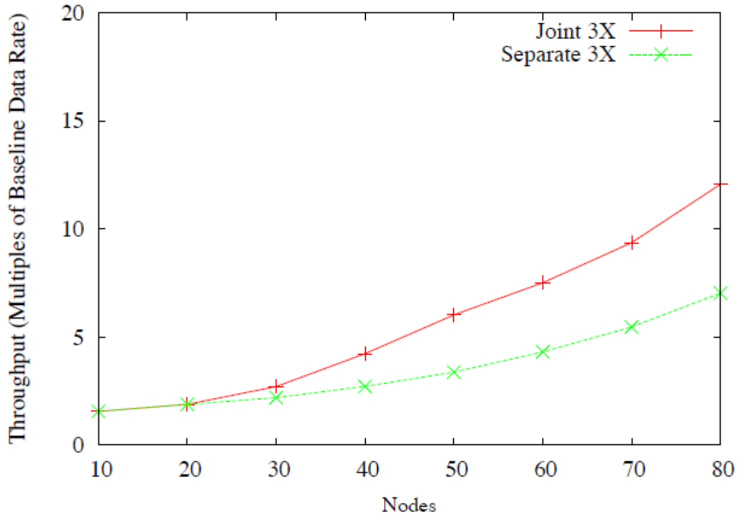
Fig. 2. Two simple test cases

To show the advantage of joint spatial-temporal multiplexing, we run simulations on more complex network topologies with multiple data flows. We assume a random network, with random (source, destination) pairs. Network sizes range from 10 to 80 nodes, among which 20% nodes are sources, each with its own destination that is randomly chosen. We observed that as the number of antennas on each node increases, the network throughput increases. The joint design starts to show performance gain after certain point, and show more performance gain in larger network sizes. Fig. 3 shows the performance comparison between the joint design and separate design given by the model in (6)-(7d).

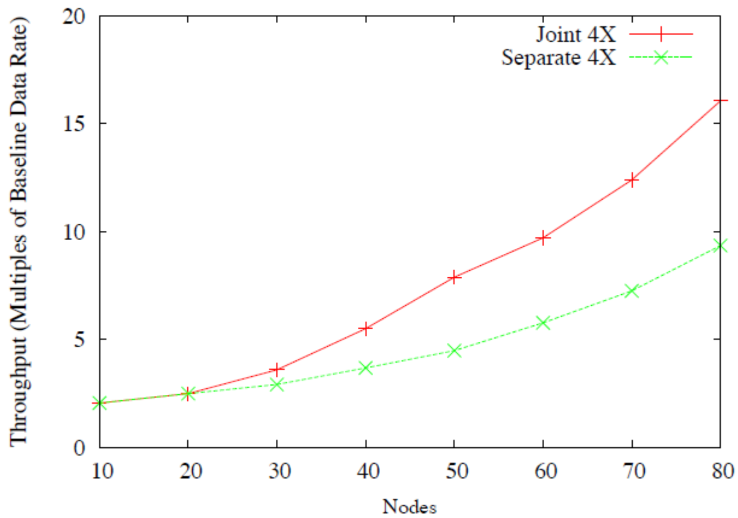


(a)

Fig. 3. Throughput as multiples of baseline data rate



(b)



(c)

Fig. 3. Throughput as multiples of baseline data rate (cont.)

(a) with 2 transmitantennas and 2 receive antennas

(b) with 3 transmit antennasand 3 receive antennas

(c) with 4 transmit antennas and 4receive antennas

V. CONCLUSIONS

In this paper, we study the multiplexing problem in a MIMO-based ad hoc network. By jointly considering spatial multiplexing and temporal multiplexing in one optimization framework, we can maximize the network throughput. A linear programming based algorithm is proposed. Simulation results verified the advantage of using such a joint optimization approach.

This paper addressed the issue of spatialtemporal multiplexing when the routing information is given. In fact, routing and spatial multiplexing can be jointly designed to further take advantage of the MIMO technology. A cross-layer design scheme would be appropriate to jointly consider cooperative routing and spatial multiplexing. This will be studied in our future work.

REFERENCES

[1] R. G. Gallager, *Information Theory and Reliable Communication*. Wiley, New York, NY, 1968.

[2] S. Toumpis and A. J. Goldsmith, “Capacity regions for wireless ad hoc networks,” *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*, vol. 2, no. 4, pp.736–748, July 2003.

[3] S. Ye and R. S. Blum, “Optimized signaling for mimo interference systems with feedback,” *IEEE Trans. Signal Processing*, vol. 51, no. 11, pp. 2839–2848, Nov. 2003.

[4] S. Cartreux, L. J. Greenstein, and P. F. Dressen, “Simulation results for an interference-limited multiple-input multiple-output cellular system,” *IEEE Commun. Lett.*,vol. 4, no. 11, pp. 334–336, Nov. 2000.

[5] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, “Iterative water-filling for gaussian vector multiple-access channels,” *IEEE Trans. Inform. Theory*, vol. 50, no. 1, pp.145–152, Jan. 2004.

[6] J. Liu, Y. T. Hou, Y. Shi, and H. D. Sherali, “Optimization of multiuser mimo networks with interference,” in *GLOBECOM’2006*, 2006.

[7] W. Ajib and D. Haccoun, “An overview of scheduling algorithms in mimo-based fourth-generation wireless systems,” *IEEE Netw.*, vol. 19, no. 5, pp. 43–48, Sep.–Oct. 2005.

[8] V. K. N. Lau, Y. Liu, and T. A. Chen, “Optimal multiuser space time scheduling for wireless communications,”in *IEEE VTC 2002-Fall*, 2002, pp. 1939–1942.

[9] D. Aktas and H. E. Gamal, “Multiuser scheduling for mimo wireless systems,” in *IEEE VTC 2003-Fall*, 2003, pp. 1743–1747.

[10] S. Chu and X. Wang, “Opportunistic and cooperative spatial multiplexing in mimo ad hoc networks,” *IEEE/ACM TRANSACTIONS ON NETWORKING*, vol. 18, no. 5, pp. 1610–1623, 2010.

[11] ——, “Mimo-aware routing in wireless mesh networks,” in *IEEE INFOCOM 2010*, 2010.

[12] M. Hu and J. Zhang, “Mimo ad hoc networks: Medium access control, saturation throughput, and optimal hop distance,” *J. Commun. Netw., Special Issue on Mobile Ad Hoc Networks*, vol. 6, no. 4, pp. 317–330, 2004.

[13] D. Qian, D. Zheng, J. Zhang, and S. N, “Csma-based distributed scheduling in multi-hop mimo networks under sinr model,” in *INFOCOM*, 2010.

[14] G. Brar, D. Blough, and P. Santi, “Computationallyefficient scheduling with the physical interference model for throughput improvement in wireless mesh networks,” in *in MobiCom, ACM*, 2006, pp. 2–13.

[15] J. Liu, Y. T. Hou, Y. Shi, H. D. Sherali, and S. Kompella, “On the capacity of multiuser mimo networks with interference,” *IEEE Transactions on Wireless Communications*, vol. 7, pp. 488–494, 2008.

[16] J. Liu, Y. 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.*, no. 6, pp. 913–926, 2008.

[17] M. Levorato, S. Tomasin, P. Casari, and M. Zorzi, “Analysis of spatial multiplexing for cross-layer design of mimo ad hoc networks,” in *IEEE VTC 2006Spring*, May 2006, pp. 1146–1150.

[18] M. Z. et al., “Cross-layer issues in mac protocol design for mimo ad hoc networks,” *IEEE Wireless Commun. Mag.*, vol. 13, no. 4, pp. 62–76, Aug. 2006.

[19] D. Tse and P.Viswanath, *Fundamentals ofWireless Communication. Cambridge*. U.K.: Cambridge Univ. Press, 2005.

[20] B. Vucetic and J. Yuan, *Space-Time Coding*. NewYork: Wiley, 2003.

V. RATE-ADAPTIVE CONCURRENT TRANSMISSION SCHEDULING IN 60-GHZ MM-WAVE WPANS

Maggie Cheng, Quanmin Ye and Lin Cai

Abstract — We consider the scheduling problem in a millimeterwave wireless personal area network in which users can use adaptive coding and modulation schemes to change their data rates. The scheduling problem is to map transmissions to time slots so that the total throughput is maximized. The challenge is that the achieved data rate of one flow is limited by the interference from other transmissions in the same slot, which is unknown until the scheduling decision is known. We propose to discretize data rate into several distinct levels supported by the PHY layer, and then use a linear programming model to find the highest rate level a flow can achieve. The same model is extended to consider a mixture of omni-directional antennas and directional antennas with heterogeneous transmitting power. The simulation results show that the proposed algorithms outperform the previous work for adaptive-rate scheduling in both throughput and fairness.

I. INTRODUCTION

Wireless personal area networks (WPANs) feature multiusers coexisting in a small area. Due to the short distance, users can communicate with each other directly in a peerto-peer fashion without involving a relay node. Also due to the short distance, nodes may have harmful interference to each other. However, from a system point of view, having the nodes access the channel one at a time in a serial TDMA manner may be a waste of resource due to the long waiting time each user experiences. The total network throughput may be decreased for not utilizing spectrum spatial reuse. The performance degradation can be significant when it comes to the mm-Wave based WPANs, since at the 60-GHz band oxygen absorption peaks and the transmission range is short, so there is large room for spectrum spatial reuse. To make efficient use of the radio spectrum, a rigorous treatment of the subject is deemed necessary. This motivates the study of finding an optimal schedule for the maximum network throughput.

Although scheduling is a well-studied subject, to optimally schedule transmissions when each user can adjust its data rate according to the signal to interference and noise ratio (SINR) is relatively new. A new dose of challenge is added when the data rate of each user is not known a priori. The achievable data rate is dependent on who else is transmitting, which is not known until the scheduling decision is made.

The rate-adaptive scheduling problem is different from the scheduling problems studied in previous work [1], [2], in which fixed data rates were used for individual flows. The main challenge is to manage the multi-user interference. However, there is no straightforward solution for interference management. How to optimize the scheduling solution for rateadaptive networks remains an open issue. In this paper, we take an optimization-based approach to determine the concurrenttransmission schedule; moreover, we deal with the interference relation in a continuous scale for maximum performance gain, which is superior to the previous work that used a binary conflict relation-based approach [3]. The proposed solution also considers fairness among users and updates the schedule at the end of each slot, which allows real-time application since the slot duration is long enough to accommodate a schedule update.

The rest of the paper is organized as follows. In Section II, we describe the network model for which the scheduling problem is studied. In Section III, we formally introduce the rate-adaptive scheduling problem. In Section IV, we proposea linear programming model to find the optimal schedule for networks with homogeneous and omni-directional antennas. In Section V, we extend the homogeneous model to networks with heterogeneous transmitting power and different antenna types. We compare the proposed work with previous related work and present simulation results in Section VI. In Section VII, we briefly survey the previous related work. Section VIII concludes the paper and points out future research directions.

II. SYSTEM MODEL

We consider the model of IEEE 802.15.3 for ultra-wideband WPANs. Nodes are deployed in a small region, typically no larger than 10m×10m. Each transmitter can directly transmit to its receiver without using a relay, and this communication pair forms a flow. One of the nodes will be selected as the coordinator of the entire piconet. Communication within the piconet starts from devices sending requests to the coordinator using a contention-based protocol, then the coordinator will make a scheduling decision and announce it to all active devices. Peer-to-peer communication among devices takes place only in the allocated time period.

Suppose there are N active flows. We name the transmitter of flow i transmitter i and the receiver of flow i receiver i. If we assign one flow per slot and arrange transmissions in a round-robin fashion, there will be no multi-user interference. A flow will achieve its highest data rate in the allocated slot, but will have to refrain from access in other slots, so the average data rate of flow i is only R_i/N, where R_i is its achieved data rate in a transmitting slot. If we use CDMA in each slot, and allow multiple flows to transmit in one slot using different codes, it is possible that the total throughput is higher. We assume the codes are pseudorandom sequences, which have a number of good properties including immune to noise and auto-correlation, and low requirement for synchronization, etc.

In this paper, we adopt the same physical layer model used in [3], which uses the 60 GHz mm-Wave unlicensed band, and uses DS-CDMA in each allocated slot. We focus on the scheduling scheme used by the coordinator for slot allocation.

III. SCHEDULING PROBLEM FOR RATE-ADAPTIVE WIRELESS NETWORKS

Scheduling rate-adaptive flows is fundamentally different from scheduling fixed-rate flows. The latter is easier in that as long as the SINR is above a threshold, it uses a fixed data rate to transmit, while in the former a transmitter varies its data rate according to the received SINR. When the ratio is high, a node increases its data rate to maximize spectrum utilization. If the transmitting power is P_T, then the received power $P_R = \kappa_1 G_T G_R P_T d^{-\gamma}$, where $\kappa_1 \propto \left(\frac{\lambda}{4\pi}\right)^{\gamma}$ is the constant scaling factor corresponding to the reference path-loss, G_T and G_R are transmit and receive antenna gains respectively, d is the distance between the transmitter and receiver, and γ is the path loss component, usually between 2 and 6. The noise and interference consists of white Gaussian noise and interference, I, from other transmitters. If the white Gaussian noise spectral density is N_0, then the total noise power is $N_0 W$. Using the Shannon's theory, the achievable data rate is

$$R = \kappa_2 W log_2 \left(1 + \frac{\kappa_1 G_T G_R P_T d^{-\gamma}}{N_0 W + I}\right)$$

where κ_2 accounts for the efficiency of the transceiver design, and W is the channel bandwidth. When there is only one active flow in the network, the achievable data rate is fully determined by its own transmitting power and constant parameters. However when multiple users share one slot, the achievable data rate for each user is no longer a constant. A transmitter then adapts its data rate based on the current SINR. It is considered *rate-adaptive* because R varies with I.

Definition 1 (MTS): The Maximum Throughput Scheduling problem is to find an optimal assignment of flows to slots such that the network throughput is maximized.

For instance, assume that flow 1 and flow 2 have data rates R_1 and R_2, respectively, when they each transmit alone. If flow 1 and flow 2 transmit at the same time, the data rate of flow 1 becomes R_1^2 with the superscript indicating the interfering flow. Apparently $R_1^2 < R_1$, and $R_2^1 < R_2$. The combined throughput is $R_1^2 + R_2^1$. The optimal solution to the maximum throughput is $\max\{R_1^2 + R_2^1, R_1, R_2\}$. With more flows involved, the optimal solution is selected from a large number of choices exponential to the number of flows. For N flows, the cardinality of the candidate solution set is $\sum_{k=1}^{N}\binom{n}{k} = 2^N - 1$. Among the $2^N - 1$ options, there is an optimal operating point that provides the maximum total throughput. However, which flows can be put in one slot to maximize the total throughput is a complicated combinatorial optimization problem. We show that the MTS problem is NP-hard in the Appendix. Apparently, an exhaustive search algorithm is not practical for large N. We henceforth explore a linear programming-based approach.

IV. LINEAR PROGRAMMING-BASED TRANSMISSION SCHEDULING

The proposed method involves first discretizing data rate to H levels, with each level corresponding to a PHY layer coding and modulation scheme. The lowest data rate, r_1, is defined as the minimum data rate at which a node is allowed to transmit. In other words, if a node cannot achieve this data rate due to interference, it won't be transmitting. The highest data rate, r_H, is the maximum data rate a node can achieve when transmitting at the maximum power to a receiver at distance d = 1and there is no interference from other flows. The actual data rate of a flow in a transmitting slot is between the two boundaries depending on the interference it receives. Variables and constants used in the linear program are listed as follows.

Variables:

- 0-1 integer variable u_i = 1 if flow i uses the current slot to transmit; =0 otherwise.
- 0-1 integer variable $t_{i,h}$= 1 if flow i uses the current slot to transmit at rate level h; =0 otherwise.
- Real-valued variable R_i is the achieved data rate of flow i in the current slot. If flow i is not transmitting, R_i = 0.

Constants:

- N is the number of flows.
- K is the number of slots.
- W is spectrum bandwidth in MHz.
- γ is the pass loss exponent.

- κ_1 is a constant dependent on wavelength, $\kappa_1 \propto \left(\frac{\lambda}{4\pi}\right)^{\gamma}$
- κ_2 is the coefficient describing the efficiency of the transceiver design.
- r_h for $h = 1..H$ are the discretized data rates. r_1 is the the minimum rate, and r_H is the maximum rate. H is the number of levels.
- $SINR_h$ is the signal to interference and noise ratio threshold to achieve data rate rh. Using equation $r_h = \kappa_2 W log_2(1 + SINR_h)$, $SINR_h$ can be calculated as follows: $SINR_h = 2^{r_h/(\kappa_2 W)} - 1$ for a given data rate r_h.
- $G_T(i)$ is the omni-directional antenna gain of transmitter i, $G_R(i)$ is the omni-directional antenna gain of receiver i. For homogeneous model, $G_T(i)$ = $G_R(i)$ = ŋ ×1, where ŋ is the antenna radiation efficiency.
- b is cross-correlation between two concurrent transmissions in a CDMA context, also called multiuser interference (MUI) factor.
- N_∞ is a very large positive number.

A. A High Throughput Scheduling Algoritthm (LP)

We propose an efficient algorithm that schedules flows for one slot at a time and sequentially apply the method to the next slot until all slots have been scheduled.

Let k be the current slot number. $R_i^{(k)}$ is the sum of data rates from slots 1 to k, for $k = 1, \ldots, K$. Upon termination, we get $R_i^{(k)}/k$, the average data rate of flow i over K slots. Initially we set $k = 1$, and $R_i^{(0)}$ = 0, $\forall i$.

Algorithm LP

- **Step 1:** Pick the flow that has the lowest throughput:

$$i^* = arg\ \min_i R_i^{(k-1)}$$

  then set u_{i^*} = 1. For the first iteration, since all flows have $R_i^{(0)}$ = 0, pick a flow randomly.

- **Step 2:** Solve the following linear program, and obtain data rate R_i and slot assignment u_i at the current slot:

  Maximize

$$\sum_{i=1}^{N} R_i \tag{1}$$

  Subject to

$$u_{i^*} = 1 \tag{2a}$$

$$u_i \leq \sum_{h=1}^{H} t_{i,h}\,, \forall i \tag{2b}$$

$$u_i \geq t_{i,h}\,, \forall i, \forall h \tag{2c}$$

$$\sum_{h=1}^{H} t_{i,h} \leq 1\,, \forall i \tag{2d}$$

$$R_i \leq \sum_{h=1}^{H} t_{i,h}\, r_h\,, \forall i \tag{2e}$$

$$N_\infty\left(1 - t_{i,h}\right) + \left(\kappa_1 G_{i,i} P_T d_{i,i}^{-\gamma} / SINR_h\right) \geq$$

$$N_0 W + b \sum_{l \neq i} \left(u_l \kappa_1 G_T(l) G_R(i) P_T d_{l,i}^{-\gamma}\right)\ , \forall i, \forall h \tag{2f}$$

- **Step 3:** update $R_i^{(k)}$:

$$R_i^{(k)} = R_i^{(k-1)} +\ R_i, \forall i$$

- **Step 4:** $k = k + 1$. While $k \leq K$, repeat steps 1–3.
- **Return:** the average data rate of each flow $R_i^{(k)}/K$ and network throughput $\sum_{i=1}^{N} R_i^{(k)}/K$.

Inequality (2b) requires that if $t_{i,h} = 0$ for all h, then $u_i = 0$, which means if flow i is not using any valid data rate to transmit, then flow i is not transmitting in slot k at all. Inequality (2c) requires that if $t_{i,h} = 1$ for some h, then $u_i = 1$, which states that if flow i is using some data rate r_h to transmit, then flow i is transmitting in slot k. Inequality (2d) requires that a flow either use one rate level to transmit or not transmit at all. Inequality (2e) states that if flow i is using level h then the achieved data rate is r_h in this slot. (2f) requires that if flow i has data rate r_h in slot k, then the signal to interference and noise ratio must be at least $SINR_h$; otherwise if $t_{i,h} = 0$, the inequality is automatically satisfied for having a large positive constant at the left hand side.

Formulating the MTS problem into a Mixed Integer Program (MIP) does not make the original problem easier to solve, since it is still an NP-hard problem, but it does provide a way to approach the optimal solution. The MIP can be solved by first relaxing it to a real-valued linear program and then rounding fractions to integers. The heuristics yields a feasible schedule, from which the actual data rates are calculated according to the Shannon's theory. Most solvers (i.e., *lp_solve*) have built-in ability to do LP-relaxation and rounding. It takes only $O(n^3)$-time to solve a real-valued linear program with n variables, much lower than the exponential-time exhaustive search algorithm.

B. Throughput-Fairness Tradeoff Design (LP-Fair)

Although the algorithm LP does have some control over fairness by allowing the flow with the lowest data rate to transmit in the next slot, the achieved data rate in the next slot still depends on the interference from other transmitters. To improve fairness, we introduce a coefficient $c_i (c_i > 0)$ for each flow i and use the following objective function in Step 2:

Maximize

$$\sum_{i=1}^{N} c_i R_i \tag{3}$$

In the first slot $(k = 1)$, we set $c_i = 1, \forall i$. When ≥ 2 , c_i reflects the current data rate of flow i. If $R_i^{(k-1)}$ is the smallest compared to other flows' data rates, then $c_i$will be the largest so that increasing R_i in the next slot will most effectively increase the objective value. If another flow, j, is interfering with flow i but has a smaller coefficient, then flow i will have a higher priority to transmit, and flow j may not be allowed to transmit if its interference decreases R_i too much. To achieve this effect, we can sort $R_i^{(k-1)}$ in a non-decreasing order and then assign c_i as the reverse order of $R_i^{(k-1)}$, thus the flow with a small $R_i^{(k-1)}$ will have a large c_i. We call this algorithm LP-Fair.

C. A Benchmark for Maximum Throughput

To assess the effectiveness of the proposed algorithm, we present a benchmark model. The benchmark model is a global optimization model that tries to schedule transmissions for all slots in one linear program, with an objective of maximizing the total throughput. All variables need to add one more dimension k to indicate slot number.

Variables:

- 0-1 integer variable $u_{k,i}$ = 1 if flow i uses slot k to transmit; =0 otherwise.
- 0-1 integer variable $t_{k,i,h}$= 1 if flow i uses slot k to transmit at rate level h; =0 otherwise.
- Real-valued variable $R_{k,i}$ is the achieved data rate of flow i in slot k. If flow i is not transmitting, $R_{k,i}$ = 0.

Maximize

$$\frac{1}{K}\sum_{i=1}^{N}\sum_{k=1}^{K} R_{k,i} \tag{4}$$

Subject to

$$u_{k,i} \leq \sum_{h=1}^{H} t_{k,i,h}\,, \forall k\,, \forall i \tag{5a}$$

$$u_{k,i} \geq t_{k,i,h}\,, \forall k, \forall i, \forall h \tag{5b}$$

$$\sum_{h} t_{k,i,h} \leq 1, \forall k, \forall i \tag{5c}$$

$$R_{k,i} \leq \sum_{h=1}^{H} t_{k,i,h}\, r_h\,, \forall k, \forall i \tag{5d}$$

$$N_{\infty}\left(1 - t_{k,i,h}\right) + \left(\kappa_1 G_{i,i} P_T d_{i,i}^{-\gamma} / SINR_h\right) \geq$$
$$N_0 W + b \sum_{l \neq i}\left(u_{k,l} \kappa_1 G_T(l) G_R(i) P_T d_{l,i}^{-\gamma}\right)\,, \forall k, \forall i, \forall h \tag{5e}$$

Solving this optimization problem will lead to the slot assignment for all slots. We call this algorithm *Aggregate*.

In addition to pursuing the maximum throughput, sometimes there is a strict requirement for fairness among flows. We adopt one of the two widely used methods to address fairness requirement, and call the algorithm with fairness control *Aggregate-Fair*.

1) Each individual flow must have data rate at least R_i^{min} . We add the following constraint to address it:

$$\frac{1}{K}\sum_{k} R_{k,i} \geq R_i^{min}, \forall i \tag{5f}$$

2) Each individual flow must achieve at least p-fraction of the total throughput, where $0 \leq p \leq 1/N$ is a constant. We add the following constraint to address it:

$$\sum_{k} R_{k,i} \geq p \sum_{i}\sum_{k} R_{k,i}, \forall i \tag{5g}$$

The solution to the linear programming model defined by (4)–(5e) leads to the maximum throughput, but it cannot be efficiently solved for large networks. With the additional fairness constraint, it is more complex. In the next section, we will use it as a benchmark to evaluate the performance of the slot-by-slot methods LP and LP-Fair.

D. Comparison to the Benchmark

We show a network with ten nodes and five flows deployed on a 10m × 10m square region. Transmission power P_T is set to 10 mW. Transmitters and receivers are shown in Fig. 1. *Aggregate-Fair* algorithm is used with $p = 1/10$. Transmission schedule is shown in TABLE I. The *Aggregate* model would schedule flow 2 and flow 3 in every slot, and TDMA would schedule one flow per slot, but the slot-by-slot model *LP* would schedule flow 1, flow 2, and flow 3 in the first slot, then flow 2 and flow 4 in the second slot, and so on. Throughput result is shown in TABLE II. Comparing *LP* with *Aggregate*, *LP* has 3.1% throughput loss; Comparing *LP-Fair* with *Aggregate-Fair*, *LP-Fair* has only 1.1%

throughput loss since the fairness constraint also reduces total throughput for *Aggregate-Fair*.

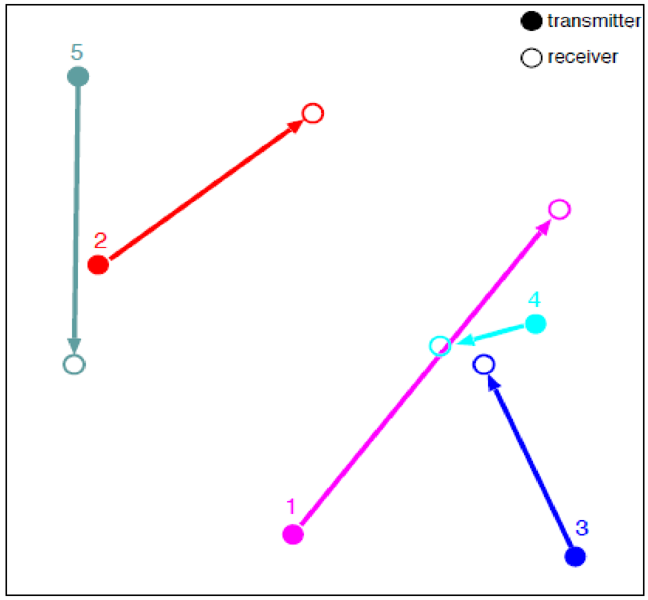


Fig. 1. A network of 10 nodes, 5 flows

Table I: Transmission scheduling

Slot	Agg	Agg-Fair	LP	LP-Fair	TDMA
1	2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1
2	2, 3	4, 5	2, 4	4, 5	2
3	2, 3	1, 2, 3	2, 3, 5	1, 2, 5	3
4	2, 3	4, 5	1, 2, 3	2, 3, 5	4
5	2, 3	1, 2, 3	2, 4	2, 4	5

Table II: Flow data rate and network throughput (Mbps)

Flow	Agg	Agg-Fair	LP	LP-Fair	TDMA
1	0	1023.52	682.34	911.02	341.21
2	3706.23	1690.68	2705.09	2147.81	563.62
3	3706.23	1519.69	1646.33	1139.76	633.27
4	0	1663.33	2049.97	1856.65	1025.09
5	0	1241.39	99.41	1005.03	497.07
$\sum$	7412.46	7138.68	7183.14	7060.27	3060.26

a) *Running-time Comparison:* Compared to the benchmark models, which schedule all slots in one linear program, the proposed slot-by-slot models reduce the number of variables by a factor of K (K is the number of slots). Solving the realvalued linear program is in the order of $O(n^3)$ where n is the number of variables. The slot-by-slot models solve the smaller linear program for K times, so the overall running time is still reduced by a factor of K^2 from the benchmark model. The running-time efficiency is achieved with a slight tradeoff in throughput performance.

V. EXTENSION TO DIRECTIONAL ANTENNA AND HETEROGENEOUS TRANSMITTING POWER

Section IV presented linear program models for homogeneous networks in which both transmitters and receivers have omni-directional antennas and constant transmission power. In this section, we consider networks with directional antennas and non-uniform transmission power. Since the detailed physical layer model is changed, the interference relation will not only depend on node positions but also the beamwidth and orientation of the axes of the radiation sectors.

A directional antenna pattern consists of a main lobe with gain $G_M = \eta \frac{2\pi}{\theta}$ and beamwidth θ and side lobes with gain $G_S = (1 - \eta)\frac{2\pi}{2\pi-\theta}$ and aggregated beamwidth $2\pi - \theta$. Let $G_T(i,j)$ be the transmit antenna gain between transmitter i and receiver j; $G_R(i,j)$ is the receiver antenna gain between transmitter i and receiver j. Apparently, if transmitter i is omnidirectional, then $G_T(i,j)$ is the same as $G_T(i)$ in Section IV regardless of the location of receiver j. Similarly, if receiver j is omni-directional, then $G_R(i,j)$ = $G_R(j)$ regardless of the location of transmitter i. When transmitter i has a directional antenna, in order to achieve the maximum directivity gain, we can make receiver i align with transmitter i's main lobe axis so that $G_T(i,i) = G_M$. Similarly if receiver i has directional antenna, we can make $G_R(i,i) = G_M$.

In heterogeneous networks, in addition to using different types of antennas, nodes can also have different transmit power. Let $P_T(i)$ be the transmit power of flow i, then the received power of flow i is calculated as

$$P_R(i) = \kappa_1 G_T(i,i) G_R(i,i) P_T(i) d_{i,i}^{-\gamma}$$

The interference power from transmitter j to receiver i is

$$I_{j,i} = \kappa_1 b G_T(j,i) G_R(j,i) P_T(j) d_{j,i}^{-\gamma}$$

With the above substantiation, we can replace inequality (2f) with the following in the linear program for *LP* and *LP-Fair*.

$$N_\infty(1-t_{i,h}) + \left(\kappa_1 G_T(i,i) G_R(i,i) P_T(i) d_{i,i}^{-\gamma} / SINR_h\right) \geq$$
$$N_0 W + b \sum_{l \neq i} \left(u_l \kappa_1 G_T(l,i) G_R(l,i) P_T(l) d_{l,i}^{-\gamma}\right) \text{ },\forall i, \forall h \quad (6)$$

Case 1: Omni-directional Transmitter and Receiver

If flow i has an omni-directional transmitter and an omnidirectional receiver, then $G_T(i,i)$ = $G_T(i)$ and $G_R(i,i)$ = $G_R(i)$. For other interfering transmitters $l \neq i$,

$$G_T(l,i) = \begin{cases} G_T(l), if\ T_x\ l\ is\ omni - directional; \\ G_M, if\ T_x\ l\ is\ directional\ and\ R_x\ i\ is\ inside\ the\ main\ lobe\ of\ T_x\ l\ ; \\ G_S, if\ T_x\ l\ is\ directional\ and\ R_x\ i\ is\ not\ inside\ the\ main\ lobe\ of\ T_x\ l; \end{cases}$$

Since receiver i is omni-directioanl, $G_R(l,i)$ = $G_R(i), \forall l$.

Case 2: Directional Transmitter and Omni-directional Receiver

Transmitter i is directional with $G_T(i,i) = G_M$. Receiver i is omni-directional with $G_R(i,i)$ = $G_R(i)$. For other interfering transmitters $l \neq i$, $G_R(l,i)$ = $G_R(i)$,and we follow the same discussion from Case 1 for $G_T(l,i)$.

Case 3: Omni-directional Transmitter and Directional Receiver

Transmitter i is omni-directional with $G_T(i,i) = G_T(i)$. Receiver i is directional with $G_R(i,i)$ = G_M. For other interfering transmitters $l \neq i$, $G_T(l,i)$ follows Case 1. Since receiver i is directional, $G_R(l,i)$ will depend on whether transmitter l is located within the main lobe of the receiver i.

$$G_T(l,i) = \begin{cases} G_M, if\ T_x\ l\ is\ inside\ the\ main\ lobe\ of\ R_x\ i; \\ G_S, otherwise. \end{cases}$$

Case 4: Directional Transmitter and Receiver

Transmitter i is directional with $G_T(i,i) = G_M$. Receiver i is directional with $G_R(i,i) = G_M$. For other interfering transmitters $l \neq i$, $G_T(l,i)$ follows Case 1, and $G_R(l,i)$ follows Case 3.

VI. PERFORMANCE EVALUATION

We evaluate the proposed schemes on the number of concurrent transmissions, network throughput, and fairness through simulation. The algorithms using objective functions (1) and (3) are named *LP* and *LP-Fair*, respectively. The proposed schemes are compared with the *REX* scheme in [3] and the serial TDMA scheme using the same network setup.

The networks are set up on a 10 × 10m^2 area. A total of 2N nodes are randomly deployed in the square region, from which N nodes are randomly selected as transmitters and the remainders are designated as receivers to form N active flows. We have tested networks of different sizes with N = 5 ~ 80.

Through all simulations we have used the following settings:

- The number of slots K in a frame is the same as the number of flows N. This is not a requirement by the proposed algorithms, but a requirement of the serial TDMA, which assigns one flow per slot in a TDMA frame.

- For transmission power P_T, with omni-to-omni transmissions, we vary the transmission power from 0.4 to 165 mW, which is equivalent to having exclusive region radius r_0=0.5m to 10m in REX [3]; if directional antennas are involved, we use a constant transmission power P_T = 10mW and vary the beamwidth θ from 6^o to 90^o with increment $\Delta\theta$= 6^o.

- Data rate levels H is set to 5.
- Other parameters: $\gamma = 4$, W = 500MHz, ŋ = 0.9, $b = 10^{-2}$, N_0 = −114dBW/MHz, κ_1 = −51dB, κ_2= 1, $N_\infty = 10^5$.

A. On Concurrent Transmissions

For a serial TDMA, the number of concurrent transmissions is exactly one. For *REX*, *LP* and *LP-Fair*, the number of concurrent transmissions in each slot is in general larger than one and non-uniform, so the average over all K slots is used. If the number of transmissions in slot k is x_k, then $X = \frac{1}{K}\sum_{k=1}^{K} x_k$ is used in the plots. For each point in the plot, we run 10 randomly generated test cases and calculate the average.

In Fig. 2, we show the average number of concurrent transmissions by *LP* and *REX* with 10, 30, and 40 flows. In Fig. 2(a) the number of concurrent transmissions decreases with transmission power due to the cross-flow interference, and increases with the number of flows due to having more choices to select from. In Fig. 2(b) and (c) when directional antennas are used, we observed that concurrency decreases with beamwidth since larger beamwidth causes larger cross-flow interference.

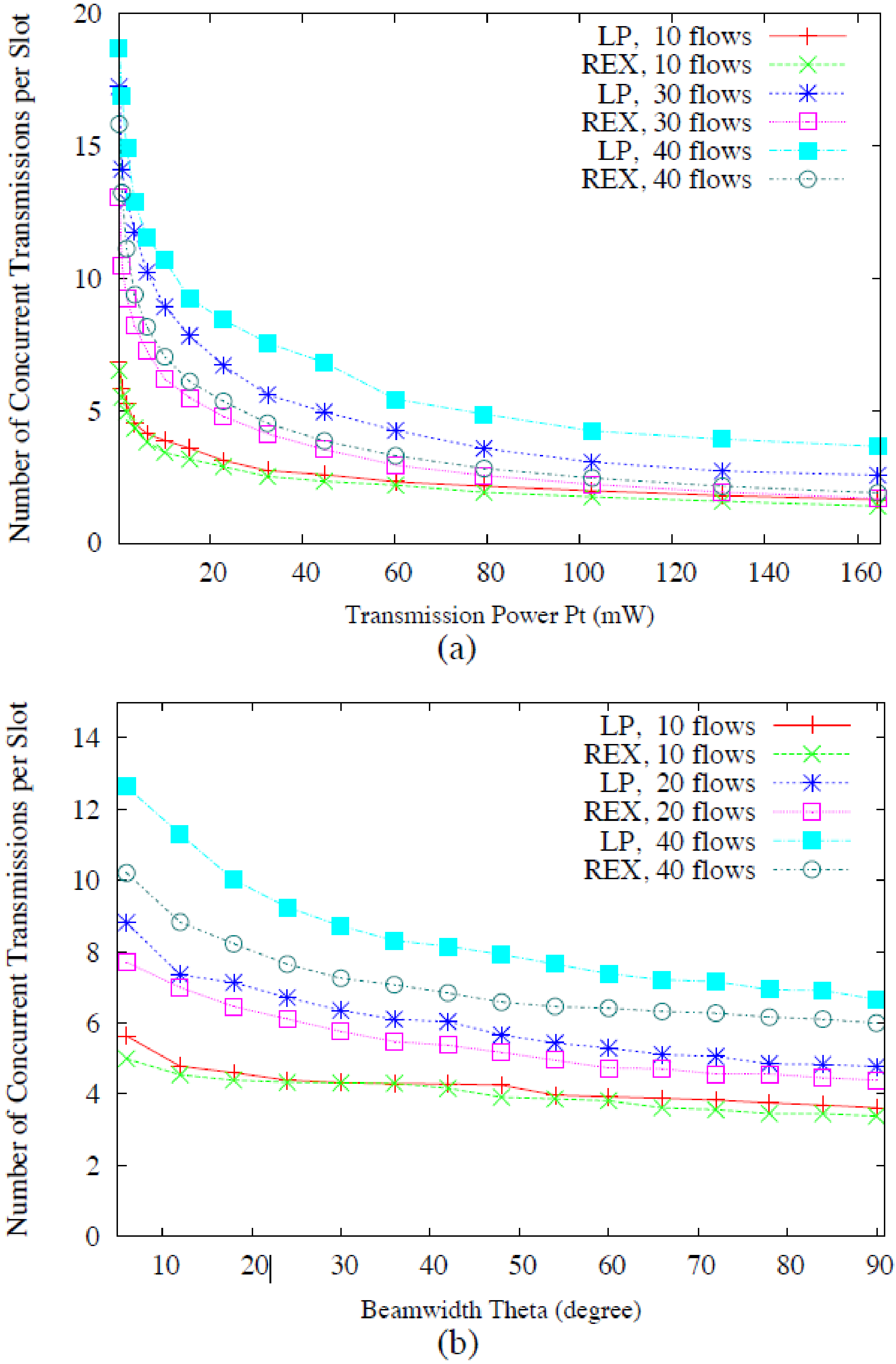
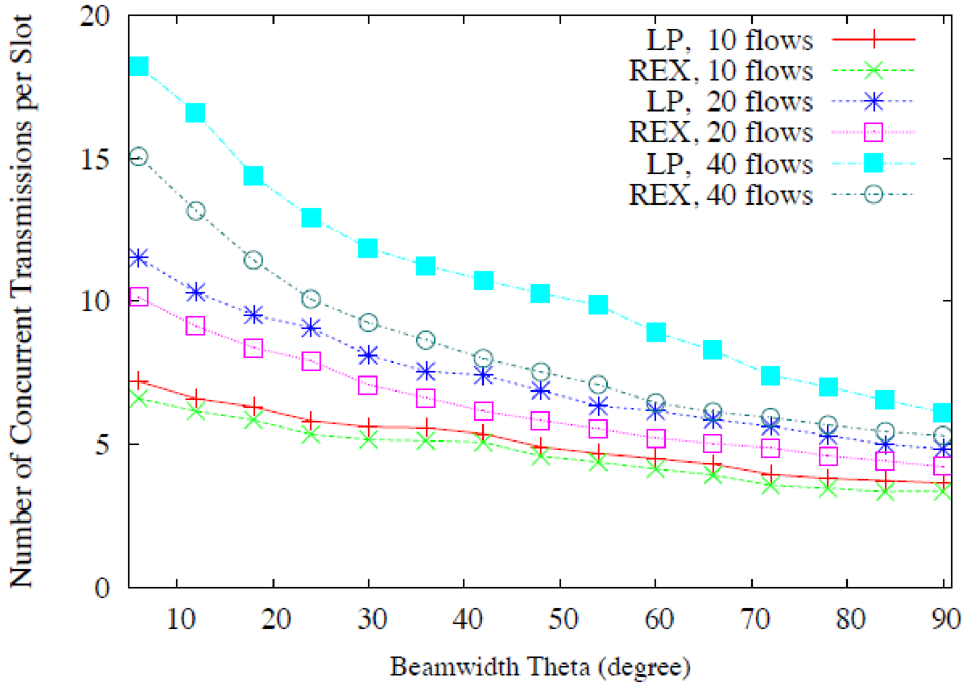


Fig. 2. Concurrent transmissions



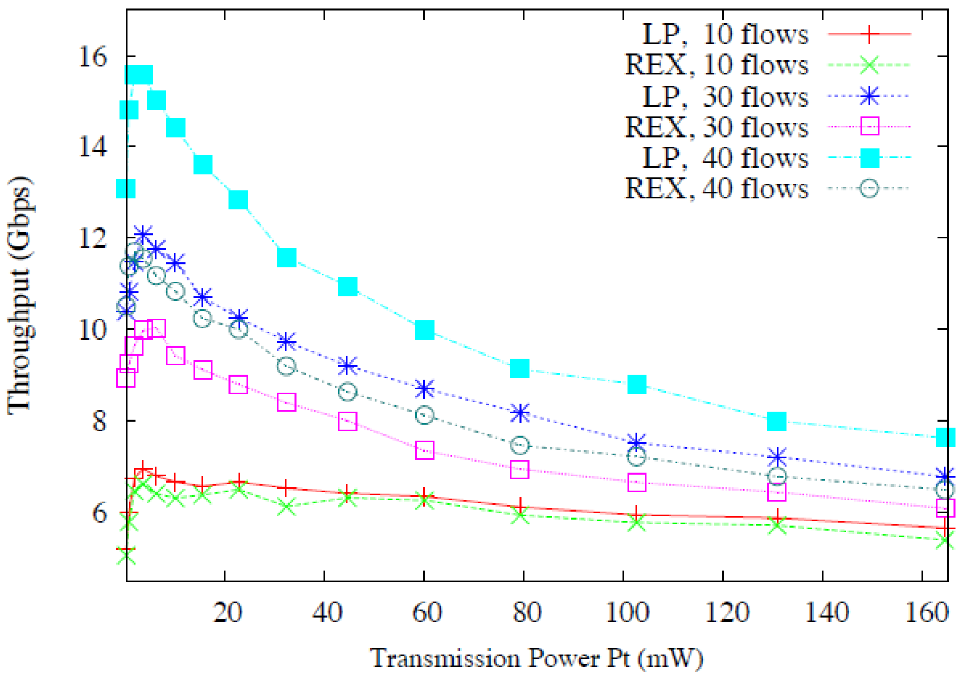
(c)

Fig. 2. Concurrent transmissions (cont.)

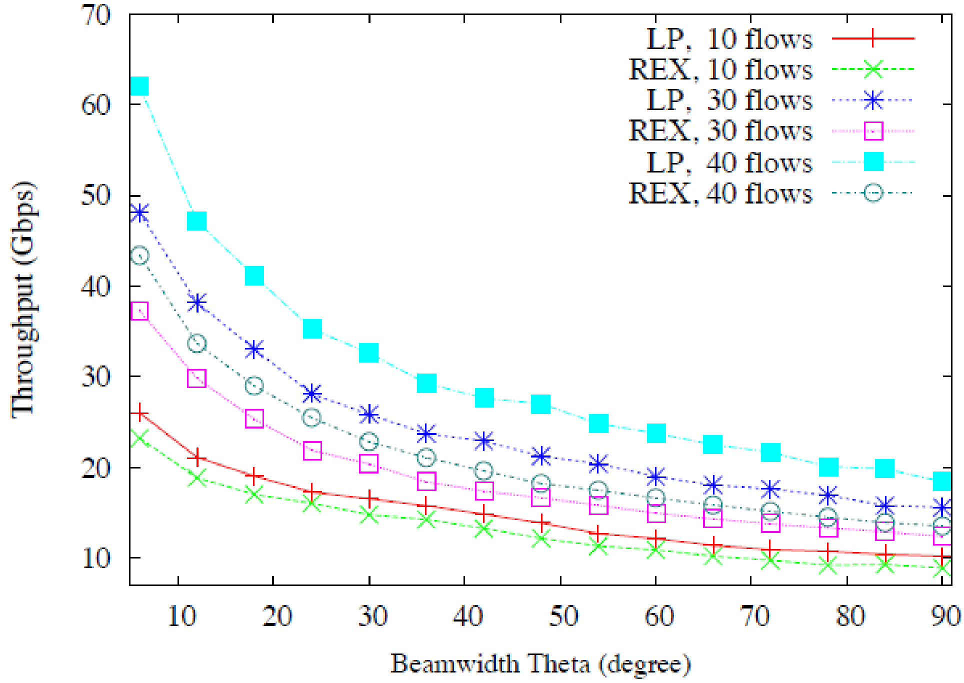
(a) Omni-Omni, (b) Omni-Directional, (c) Directional-Directional.

B. On Throughput

Fig. 3 shows network throughput achieved on the same network instances used in Fig. 2. In Fig. 3(a), as we increase the transmission power, there is an initial climbing phase from 0.4 mW to around 3.75 mW; after this point, network throughput decreases with transmission power due to stronger cross-flow interference. In Fig. 3(b) and (c), we use constant transmission power 10 mW, and network throughput decreases with beamwidth due to larger interference area.

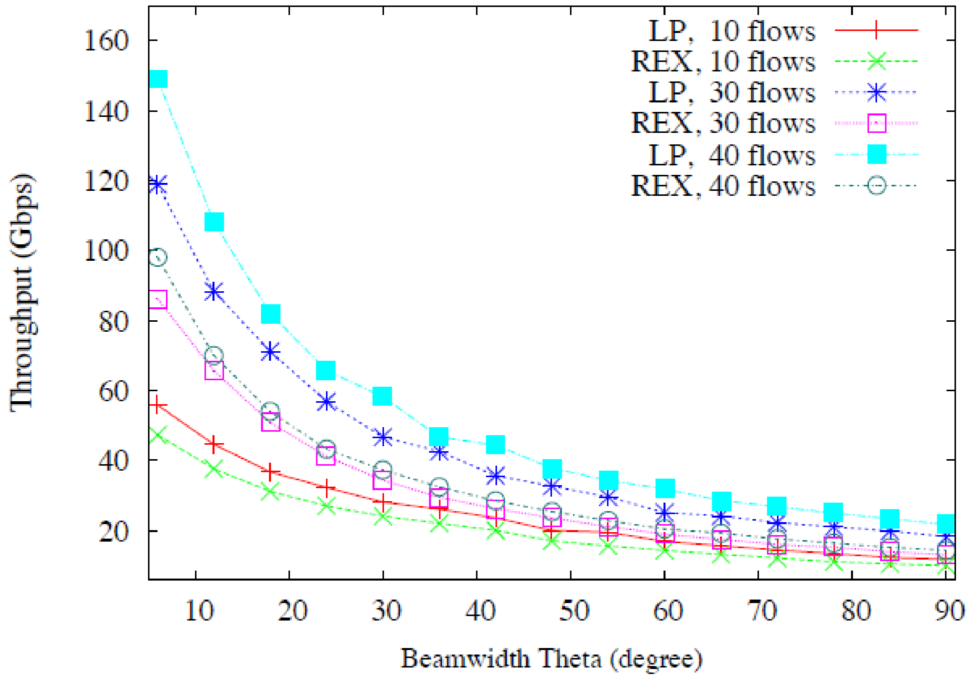


(a)



(b)

Fig. 3. Throughput

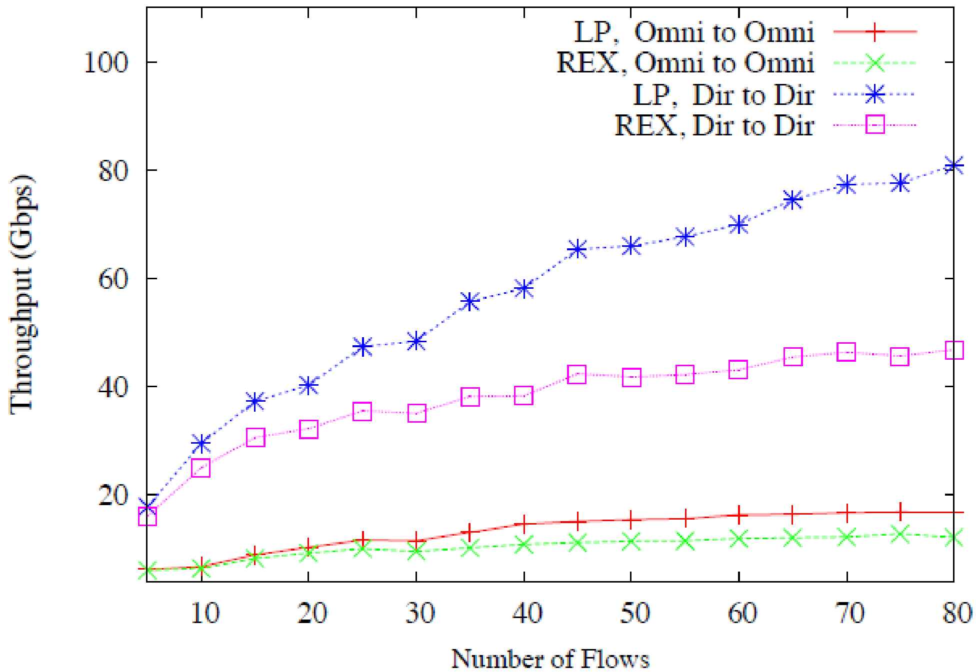


(c)

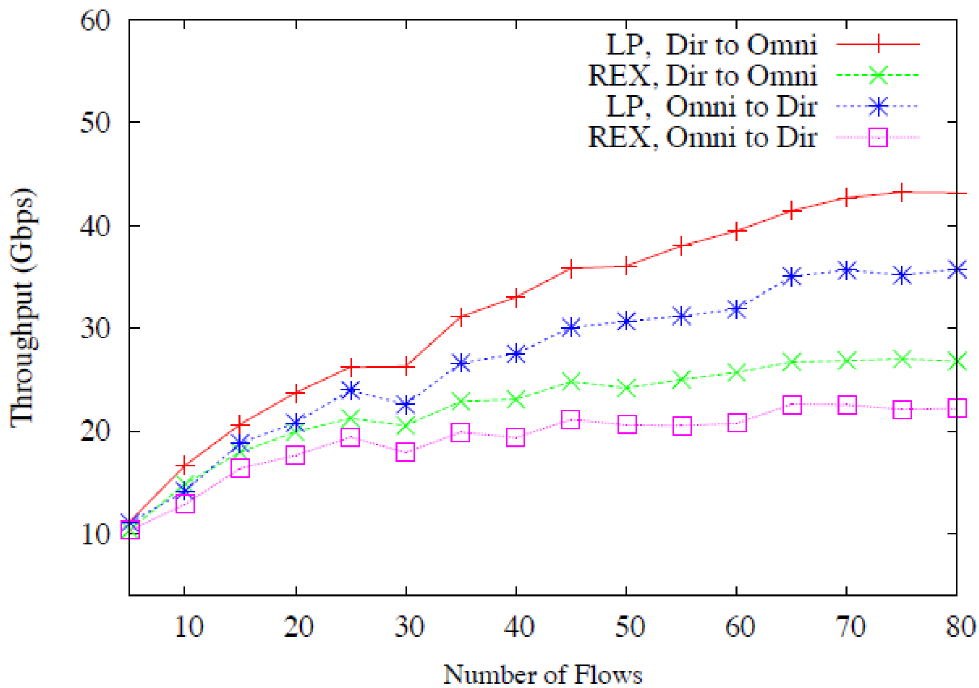
Fig. 3. Throughput (cont.)

(a) Omni-Omni, (b) Omni-Directional, (c) Directional-Directional.

We further investigated the impact of having more flows on network performance by using a fixed transmission power 10 mW and a fixed beamwidth 30^o. The results in Fig. 4 confirmed the increasing trend of throughput with the number of flows, and the conclusion is true for both omnidirectional antennas and directional antennas.



(a)



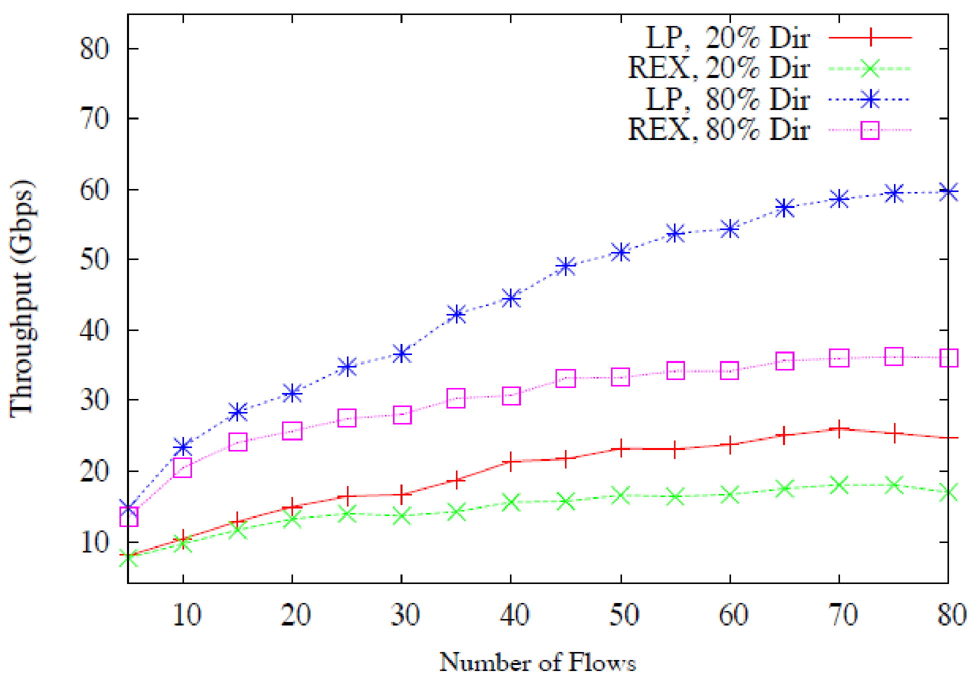
(b)

Fig. 4. Throughput

(a) Tx and Rx use the same type of antennas,

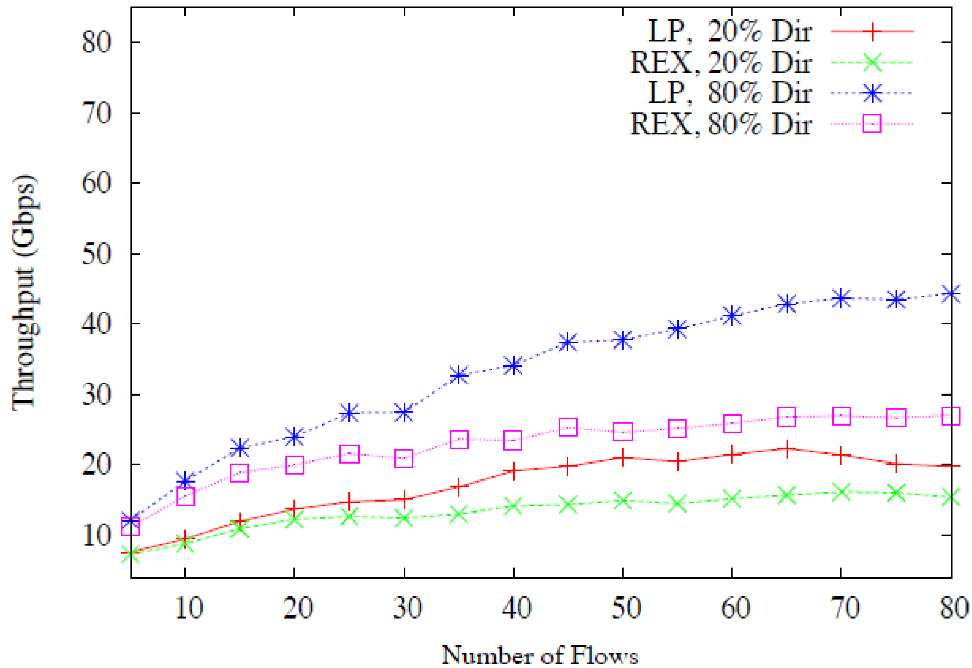
(b) T_x and R_x use the different type of antennas.

We also evaluated the algorithms in heterogeneous networks by randomly selecting either an omni-directional or a directional antenna for each node while keeping the fraction of nodes with directional antennas a fixed constant p. We ensure that if a transmitter is directional, it must face the receiver; if a receiver is directional, it must face the transmitter. We compare network throughput resulting from *LP* and *REX*. Fig. 5(a) is the result with beamwidth $\theta = 30^o$, and (b) with $\theta = 45^o$. The result shows *LP* outperforms *REX* in every single case and the throughput gain is increasing with number of flows and the percentage of directional antennas. With $\theta = 30^o$, at $p = 20\%$, the throughput gain of *LP* over *REX* is from 4% to 45%, increasing with the number of flows; and at p=80%, the throughput gain is from 9% to 65%. With $\theta = 45^o$, the throughput gain is relatively smaller and the range of throughput gain is from 4% to 28% at p=20% and from 8.5% to 64% at p=80%.



(a)

Fig. 5. Throughput result for heterogeneous networks



(b)

Fig. 5. Throughput result for heterogeneous networks (cont.)

Throughput result for heterogeneous networks with 20% or 80% directional antennas

(a) $\theta = 30^o$, (b) $\theta = 45^o$.

C. On Fairness

To evaluate peer fairness, we use Jain's fairness index. Let y_i be the measurement for flow i, then Jain's fairness is computed as $(\sum_{i=1}^{N} y_i)^2/(N\sum_{i=1}^{N} y_i^2)$. We first compare resource allocation fairness by using the number of slots allocated to each flow as measurement, then compare data rate fairness by using the per flow data rate as measurement. We call the two indices slotindex and rate-index respectively. Apparently the serial TDMA scheme has a perfect slot-index of 1.

We compare LP, LP-Fair and REX on their slot-index and rate-index. Since there exists a tradeoff between network throughput and peer fairness, we also included network throughput to show the overall performance of these schemes. We use fixed transmission power P_t = 10 for all cases, and use $\theta = 30^o$ for directional antennas. Fig. 6 shows the cases in which transmitter and receiver both use omni-directional antennas, and Fig. 7 shows the cases in which transmitter and receiver both use directional antennas. It is shown that LP-Fair has the best fairness performance and the second best throughput performance; LP has the best throughput performance, and the second best fairness performance. Both LP and LP-Fair outperform REX in throughput and fairness simultaneously.

It is further observed that every scheme has a larger slotindex than its rate-index. Since slot allocation is our means, slot-index can be easily controlled; but data rate is the outcome of using slot allocation, so rate-index can only be indirectly controlled. *LP-Fair* outperforms the other two schemes in both slot-index and rate-index, and it is noteworthy to point out that *LP-Fair* has the smallest gap between slot-index and rate-index, which indicates effective control over peer fairness.

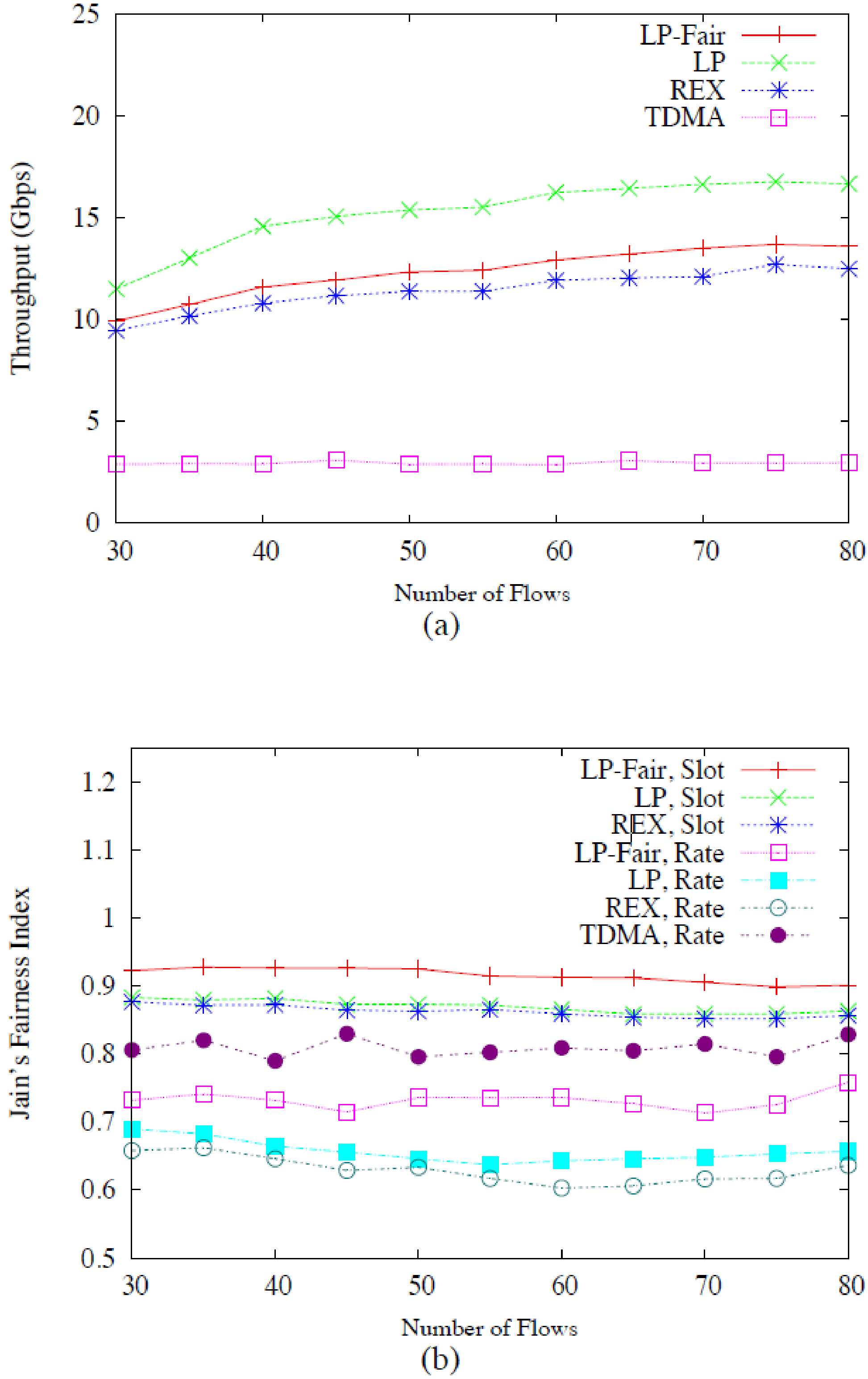


Fig. 6. Fairness result for omni-omin transmissions

(a) Network throughput, (b) Jain's index

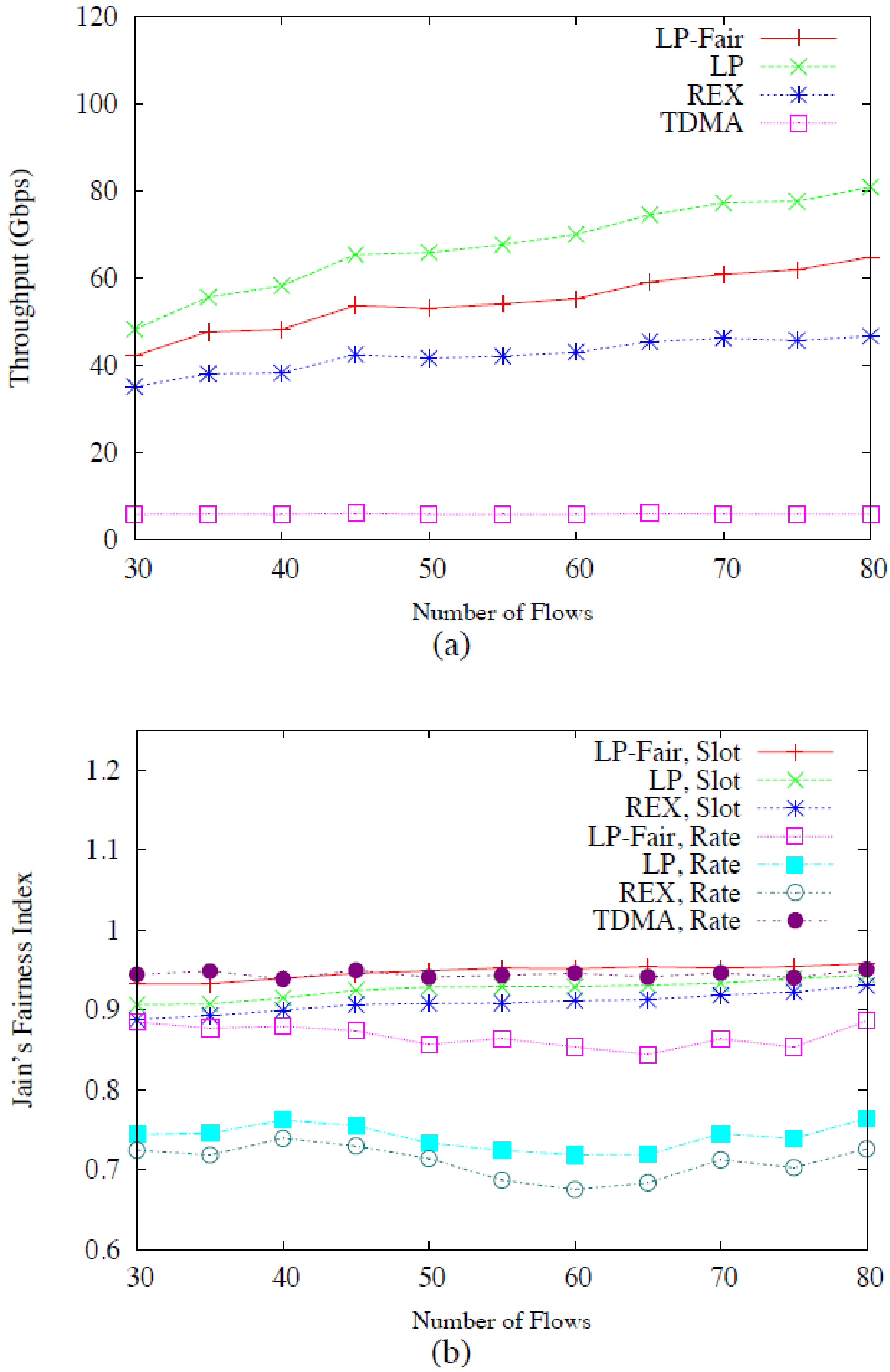


Fig. 7. Fairness result for Dir-Dir transmissions

(a) Network throughput, (b) Jain's index

VII. RELATED WORK

Scheduling is a major research problem in wireless communication and has received extensive study since the earliest days of wireless communication. Different network models have been considered. Some deal with peer-to-peer communication in personal area networks [3]–[6], some deal with one-hop communication in cellular networks [7], and some deal with multi-hop communication in ad hoc networks [1], [2], [8]. The performance consideration is mainly throughput and fairness among users.

In [9]–[11], a conflict graph is constructed to bound the mutual interference so the SINR of the tagged transmission can be above certain threshold for a predetermined data rate. As a result, fixed data rates for all transmitter-receiver pairs are used throughout the communication session. In broadband wireless systems such as UWB and mm-Wave wireless networks, the transmitter can adjust the data rate according to the SINR. With such rate-adaptive property, these conflict graph based solutions are no longer optimal. To schedule concurrent transmissions in a rate-adaptive network, [12]–[14] proposed scheduling solutions based on the exclusive region concept, which reserves an area for each flow to avoid harmful mutual interference. It is found that these concurrent scheduling can result in much higher network throughput than TDMA. [4] further proposed a global search-based algorithm to achieve higher throughput with concurrent transmissions.

Proportionally fair scheduling [15], [16] has been studied for simplified network models in [17]–[20], in which at most two users are allocated to any slot. The cardinality of the candidate solution set is polynomial, and therefore it renders a polynomial-time algorithm. Different from previous work, this paper deals with unbounded users in any slot.

The optimal solution is selected from an exponential-sized candidate set, and it is an NP-hard problem.

The most related work to this paper is [3] in which the same network model is considered, where nodes can use directional antennas and data rate can be adapted to the received SINR. The method used in [3] is to derive a sufficient condition that ensures the aggregated data rate from concurrent transmissions be higher than the would-be average data rate in a serial TDMA scheme, and then compute an exclusive region for each flow based on this sufficient condition and antenna directivity. The cheduling process starts from randomly selecting a flow, and hen adds flows one-by-one if the new flow and existing flows are out of each other's exclusive region. The benefit of this approach is its simplicity, since it is easy to determine if a node is inside a region; the drawback is that it has turned a continuous-scaled interference relation into a binary relation-ransmitters inside the region are forbidden to transmit, and those outside of the region are allowed to transmit. Thus the two nodes close to the boundary of the exclusive region with one inside and one outside are treated differently, but in fact, both have interference to the receiver and their interferences are very close in magnitude. Moreover, multiple transmitters outside of the exclusive region have accumulated effect on the receiver and their aggregated interference maybe even higher than from the one inside the exclusive region, but they could be allowed to transmit at the same time if their exclusive regions are mutually exclusive. These are the reasons for performance loss. [3] is an improvement and extension of earlier works that use the concept of exclusive region [12]–[14].

VIII. CONCLUSION

We considered scheduling concurrent transmissions in a variable data rate WPAN. Linear programming-based algorithms for maximum throughput with fairness consideration have been proposed. The simulation results showed significant improvement over TDMA and earlier work with concurrent transmissions. The performance gain increases with the directivity of the antenna and the percentage of nodes with directional antennas in the network. The proposed algorithms LP-Fair is the best in fairness and the second best in throughput, and LP is the best in throughput and the second best in fairness; the two algorithms outperform previous work REX in both fairness and total throughput.

The paper serves as a good starting point for more advanced research problems in this area, e.g., considering fast fading channels and node mobility. In addition, although the current work allows for heterogeneous transmitting power, power is used as a predetermined parameter. Future work will also consider adaptive power control and the joint design of scheduling and power control.

REFERENCES

[1] M. Cheng and Q. Ye, "Transmission scheduling based on a new conflict graph model for multicast in multihop wireless networks," in *IEEE Globecom 2012*, 2012.

[2] M. Cheng, Q. Ye, and L. Cai, "Cross-layer schemes for reducing delay in multihop wireless networks," *Wireless Communications, IEEE Transactions on*, vol. 12, no. 2, pp. 928–937, 2013.

[3] L. Cai, L. Cai, X. Shen, and J. W. Mark, "Rex: A randomized exclusive region based scheduling scheme for mmwave wpans with directional antenna," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 1, pp. 113–121, 2010.

[4] Z. Yang, L. Cai, and W.-S. Lu, "Practical scheduling algorithms for concurrent transmissions in rate-adaptive wireless networks," in *IEEE Infocom'10*, Mar. 2010, pp. 1–9.

[5] M. Bilal and M. Kang, "Throughput enhancement by concurrent transmission in wpan with multiple antennas," *in 14th International Conference on Advanced Communication Technology (ICACT),* Feb. 2012, pp. 67–71.

[6] Y. Sadi and S. Ergen, "Optimal power control, rate adaptation, and scheduling for uwb-based intravehicular wireless sensor networks," *IEEE Trans. on Vehicular Technology*, vol. 62, no. 1, pp. 219–234, Jan. 2013.

[7] L. Xu, X. Shen, and J. W. Mark, "Dynamic fair scheduling with qos constraints in multimedia w-cdma cellular networks," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 60–73, Jan. 2004.

[8] Z. Liu, M. Yang, H. Dai, and J. Dai, "Concurrent transmission scheduling for multi-hop multicast in wireless mesh networks," in *IEEE International conference WiCOM*, 2008.

[9] M. Alicherry, R. Bhatia, and L. Li., "Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks," in *In Proc. ACM Mobicom'05*, 2005, pp. 58–72.

[10] X. Li, A. Nusairat, Y. Wu, Y. Qi, J. Zhao, X. Chu, and Y. Liu., "Joint throughput optimization for wireless mesh networks," *IEEE Transactions on Mobile Computing*, vol. 8, no. 7, pp. 895–909, 2009.

[11] J. Yuan, Z. Li, W. Yu, and B. Li., "A cross-layer optimization framework for multihop multicast in wireless mesh networks," *IEEE JSAC*, vol. 24, no. 11, pp. 2092–2103, 2006.

[12] K. Liu, L. Cai, and X. Shen, "Exclusive-region based scheduling algorithms for uwb wpan," *IEEE Transactions on Wireless Communications*, vol. 7, no. 3, pp. 933–942, 2008.

[13] B. Radunovic and J. L. Boudec, "Optimal power control, scheduling, and routing in uwb networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 7, pp. 1252–1270, 2004.

[14] K. Liu, L. Cai, and X. Shen, "Multi-class utility-based scheduling for uwb networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 2, pp. 1176–1187, 2008.

[15] F. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *J. Operational Research Society*, vol. 49, no. 3, pp. 237–252, Mar. 1998.

[16] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277–1294, June 2002.

[17] X. Wang and L. Cai, "Proportional fair scheduling in hierarchical modulation aided wireless networks," *IEEE Trans. on Wireless Communications*, vol. 12, no. 4, pp. 1584–1593, April 2013.

[18] M. Hossain, M.-S. Alouini, and V. Bhargava, "Two-user opportunistic scheduling using hierarchical modulations in wireless networks with heterogenous average link gains," *IEEE Trans. Commun.*, vol. 58, no. 3, pp. 880–889, Mar. 2010.

[19] A. Jdidi and T. Chahed, "Impact of hierarchical modulation on proportional air in ofdma-based networks," in *ComNet*, 2010.

[20] A. Agustin, J. Vidal, and O. Muoz, "Performance of downlink schedulers with superposed or orthogonal transmissions," in *IEEE ICC*, 2010.

[21] M. Garey and D. Johnson, *Computers and Intractability*, *A Guide to the Theory of NP-Completeness*. FREEMAN, 1974.

APPENDIX

In this section we prove that the MTS problem is NP-hard. Apparently the MTS problem ∈ NP. The NP-hardness of MTS can be established by transforming from the Maximum Weight Independent Set problem (MWIS). The cardinality version Maximum Independent Set problem (MIS) is NP-hard [21], so is the weighted version, since the cardinality version is a subclass of the weighted version.

We first introduce a new definition.

Definition 2 (Conflict Pair): Let R_j i be the data rate of flow i when being interfered by flow j, and R_j^i be the data rate of flow j when being interfered by flow i. If flow i and flow j when sharing one slot have a combined data rate $R_i^j + R_j^i \leq max\{R_i, R_j\}$, then flow i and flow j are a Conflict Pair.

The reduction from an instance of MWIS to an instance of MTS is as follows: for any given graph $G = (V, E)$ with vertex weight set W, construct a wireless network with $2|V|$ nodes and $|V|$ flows. Each flow when not being interfered by other flows has data rate $R_i = W_i$. In addition, the inter-flow relation is configured according to the following rules:

1) If vertex i and vertex j are connected by an edge in the given graph, flow i and flow j must be a conflict pair in the wireless network.
2) Adding a new flow k to a group of flows must not change the conflict pairs of existing flows. For example, if $R_i^j + R_j^i \leq max\{R_i, R_j\}$, then $R_i^{jk} + R_j^{ik} \leq max\{R_i^k + R_j^k\}$, and if $R_i^j + R_j^i \ > max\{R_i, R_j\}$ then $R_i^{jk} + R_j^{ik} > max\{R_i^k + R_j^k\}$.

Since channel condition, transmitting power, antenna characteristics, and node location can be arbitrarily set, we can ensure that the constructed wireless network comply with the above two rules. Thus each instance of the MWIS problem is transformed to an instance of the MTS problem in polynomial time. The optimal solution to the MTS problem is a subset of flows that has the maximum combined data rate, corresponding to the subset of independent vertices in the graph that has the maximum weight.

This completes the proof that MTS is NP-hard. ■

SECTION

2. CONLUSION AND FUTURE WORK

2.1 CONCLUSION

This dissertation has provided generic mathematical models for several optimization problems in wireless networks. Two important problems have been further investigated: (1) How to minimize the end-to-end delay in multihop wireless network? Two cross-layer design schemes were proposed, and a linear program based on the sufficient condition was developed. For multicast application, conflict graph model firstly built and based on the conflict model, a linear programming model was proposed to compute the schedule of relay nodes. (2) How to maximize the throughput in wireless network? Routing, spatial multiplexing and scheduling were jointly considered, a linear programming based algorithm was proposed that included three design problem in one optimization framework. For WPAN, linear programming-based algorithms with fairness and without fairness for maximum throughput were proposed.

2.2 FUTURE WORK

Although our proposed algorithms allow for heterogeneous transmitting power in wireless personal area networks, power is used as predetermined parameter. Future work will also consider adaptive power control and the joint design of scheduling and power control.

VITA

Quanmin Ye was born in China. He received his B.S. degree in Computational Mathematics and Applied software from Central South University, China, in 1999 and M.S. degree in Applied Mathematics from South China Normal University, China, in 2004. He joined Missouri University of Science and Technology (formerly the University of Missouri-Rolla) in August 2010 and received the degree of Doctor of Philosophy in Computer Science in May 2014.