Cross-Layer Multicommodity Capacity Expansion on Ad Hoc Wireless Networks of Cognitive Radios

Lex Fridman*, Steven Weber†, Kapil R. Dandekar†, Moshe Kam†* *
Department of Computer Science
†Department of Electrical and Computer Engineering
Drexel University, Philadelphia, Pennsylvania 19104

***Abstract*— Cognitive radios permit dynamic control of physical layer resources such as transmission power and constellation size; these degrees of freedom can be employed to achieve significant improvements in network throughput above that obtainable using conventional radios (with fixed transmission power and constellation size). In this paper we present a unified framework for coordinated resource allocation across the entire protocol stack: physical, medium access, network, and transport layers. Our focus is on demonstrating that joint optimization over transmission power, constellation size, scheduling, and multicommodity flow can achieve greater network performance over optimizing resource allocation at each layer individually. We present three cases where a modularized network design problem can be "merged" and then characterize the benefit achieved by the merger.**

I. Introduction

Cross layer design has received enormous attention in recent years due to the increasing awareness that conventional network layering (ubiquitous in wired networks) unnecessarily limits achievable network capacity due to the inherent performance dependencies across layers. In this work we show three cases in which integration of design across layers provides improvement over the conventional layered design framework. The metric for performance is global throughput which is optimized by expanding link capacity with the free variables coming from a simplified version of the network stack consisting of three primary components: i) a physical layer model consisting of cognitive radios, ii) a medium access layer model where scheduling is performed, and iii) a network / transport layer model using multicommodity flow. We next describe each of the three layers and their resources.

A. Physical (PHY) layer: cognitive radios

Cognitive radios [1] permit dynamic control of fundamental wireless network physical layer parameters such as transmission power and constellation size. This dynamic control promises significant performance improvements above conventional radios with static resource allocation configurations. In this work we study the performance impact of cognitive radios that dynamically adjust their transmission power and constellation size in response to channel and interference state, in order to maximize network throughput (defined below) while maintaining an acceptably low bit error rate (BER).

PHY resource: constellation size. We assume that the cognitive radios make use of tunable m-ary Quadrature Amplitude Modulation (m-QAM) with a constellation size of m. A transmitting node achieves a throughput of $c = \log_2 m$ bits per transmission, and an associated BER that is proportional to m and inversely proportional to the receiver's signal to interference plus noise ratio (SINR). Each transmitting node seeks to find the largest constellation size that can be supported by channel and interference conditions and does not violate a specified quality of service constraint on BER.

PHY resource: transmission power. Besides adjusting the constellation size, cognitive radios are also presumed to have control over their transmission power. It is instructive to emphasize how these two degrees of freedom may be used in a complementary fashion. Adjusting the constellation size, m, increases the link capacity, $c = \log_2 m$, but also increases the bit error rate (since more points are added the constellation, thereby pushing the points closer together and increasing the chance of a symbol decoding error). Changing m has no effect on the neighboring nodes, other than to change the link capacity to the node's receivers. Adjusting the transmission power, p, on the other hand, has no effect on the link capacity. Increasing p will increase the SINR seen by the intended receiver associated with the transmitter, but will negatively impact the SINR of all other receivers.

B. Medium access control (MAC) layer

It is well known that local transmission scheduling (whether in time, frequency, or code) is an essential component to achieving good network capacity. The problem of scheduling in wireless networks is essentially a packing problem: concurrent transmissions must be sufficiently separated in space so that their respective signals do not cause undue interference on their respective receivers.

MAC resource: scheduling. Using the transmission powers and constellation sizes for each node from the physical layer, a MAC layer schedule is formed by randomly packing each time slot with a set of transmitters such that the specified BER constraint is not violated at any node designated to receive in that slot. New time slots are added until each node is selected to participate in at least one time slot.

Each time slot in the schedule can be visualized as a disconnected directed graph, with edges emanating from each transmitting node to the potential receivers associated with each transmitter (see Figure 1). By successive relaying a packet may traverse the network from any source s to any destination d provided there exists a path connecting s to d in the graph formed by the union of the individual time slot graphs.

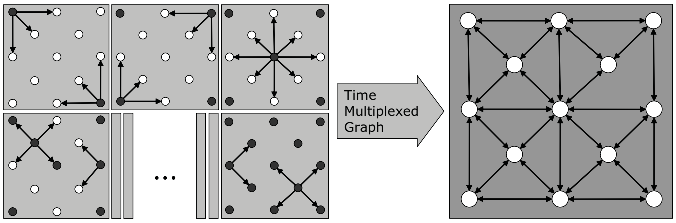

Fig. 1. Illustration of the formation of the flow graph from the underlying graphs for each slot in the schedule. The flow graph idea is adapted from [2]. The left figure illustrates the concurrent transmissions at each time slot. Black nodes are nodes that have one or more transmission slots; scheduling stops once each node is assigned to least one transmission slot. The right figure shows the flow graph obtained by time multiplexing the individual per slot graphs.

C. Network and transport (NET) layers

In the interest of parsimonious modeling we consider a fairly abstracted view of the network and transport layers using the framework of multicommodity network flow. This abstraction is justified by the observation that a multicommodity network flow problem combines the key roles of identifying *which* links should carry the traffic (the role of the routing algorithm), and *how much* data a source can afford to send on the network given the capacity constraints and congestion (the role of the transport layer).

Each commodity on the network consists of a source node and a destination node, where the source node wishes to send as much data (flow) through the network to its destination as possible (we assume infinite backlog, i.e., sources never run dry). The goal of our multicommodity network flow problem is to maximize the sum commodity rate subject to conservation of flow and capacity constraints.

NET resource: multicommodity flow. In contrast to conventional flow problems, a flow formulation for ad hoc networks must incorporate the inherent broadcast nature of wireless communications. In particular, information is multicast to all in-range receivers “for free”, but by the same token a receiver is subject to all in-range interfering nodes. This fundamental fact alters the conventional network flow problem in two significant ways.

First, all links (i, j) emanating from a given node i have a constant capacity: $c_i = \log_2 m_i$ for all $j \in \Gamma(i)$, the neighborhood of i. This differs from the conventional network flow setup where link capacities can vary across edges emanating from a node.

Second, as discussed above, the graph on which the flow is placed is the time-multiplexed union of the individual elementary capacity graphs. As such the *effective capacity* of a node is its nominal capacity, c_i, thinned by the fraction of time that the node is eligible to transmit under the specified schedule. Suppose the network transmission schedule consists of S equal duration time slots and node i is assigned to transmit on s_i of the time slots. Then the effective capacity for node i is $\tilde{c}_i = w_i \log_2 m_i$, where $w_i = s_i/S$ is the capacity thinning factor due to scheduling.

D. Contributions

Summary of model. To recap, our model of the network resources consists of three components. First, the physical layer resources are the transmission power and the constellation size, both of which are assumed to be tunable. Second, the medium access control layer provides a temporal schedule to prevent nearby nodes from simultaneous transmissions; this schedule is formed by ensuring the BER (which depends upon the transmission powers and the constellation sizes) at each receiver active in each time slot is acceptably high. Third, the resource allocation decisions of the network and transport layers are modeled by solving a multicommodity network flow problem. The solution of this problem, which depends upon the locations of the commodities sources and destinations, the network topology (determined by the schedule), and the link capacities (determined by the constellation sizes). It is clear that there are significant cross-layer resource allocation dependencies in this model.

E. Related work

Cognitive radios. Advancements in flexible and dynamic control of physical layer resources have spurred widespread interest in cognitive radios, e.g., [1], [3], [4], [5], [6], [7]. Our own work [8], [9] has focused on how ad hoc network capacity may be improved when transmitters have the channel state information (CSI) required to make intelligent resource consumption decisions, e.g., avoiding transmission during deep fades. Although existing work has dealt with many design and performance issues surrounding cognitive radios, to the best of our knowledge no work has yet addressed the problem of how to intelligently couple power control with constellation size control in a cross layer context.

Cross-layer design.

There is insufficient space here to do justice to the enormous amount of work that has gone on over the past five years or so in the field of cross layer design. We mention only the paper closest in spirit to ours, namely, the recent 2005 paper by Wu et al. [2], which is the inspiration for this work. The authors formulate a cross layer design problem across the PHY, MAC, NET layers. By solving a linear program for minimizing sum power subject to specified SINR constraints, they form a time-multiplexed schedule of *elementary capacity graphs*, which are the concurrent transmissions in each time slot. This graph is then the basis for solving a sum of max of flows problem. The most significant idea we have borrowed from this formulation is the idea of obtaining a capacitated graph for network flow problems from a time multiplexed transmission schedule. The primary differences from [2] is our focus on the value of tunable constellation sizes, and on distributed algorithms ([2] only discusses centralized algorithms).

Multicommodity flow and capacity expansion. Multicommodity flow problems are discussed at length in many textbooks on network flow, e.g., [10]. More closely related to our work in this paper, however, is the body of literature focused on capacity expansion algorithms. In contrast to the conventional network flow setting where the capacity of the underlying graph is assumed fixed, the capacity expansion problem studies the problem of joint optimization of both flow and capacity. This is a natural framework in the context of cognitive radio, since capacity is determined by the constellation size, which can be adjusted based on flow. Although several variants of the capacity expansion problem have been studied, e.g., [11], [12], none of them consider the particular case of wireless networks explicitly.

The rest of this paper is as follows. In Section II we formally define our model of the PHY, MAC, and NET layers, and state the (centralized) global cross-layer optimization problem along with the three ways to modularize this problem. Section III compares the throughput obtained through joint and layered design. The paper concludes in Section IV.

II. Problem definition

In this section we formally state our mathematical model and define the multiple-layer optimization problem in both its joint and modular forms.

A. Physical layer: power and constellation size

Consider an ad hoc network consisting of a collection of N nodes $[N] = \{1, \dots, N\}$. Define $\mathbf{x} = \{x_1, \dots, x_N\}$, where x_n is the location of node n at some snapshot in time t. Each node is equipped with a cognitive radio permitting dynamic selection of both transmission power and constellation size. Let $p_{\max}$ be the uniform per node maximum power constraint and $m_{\max}$ the maximum constellation size. Let $\mathbf{m} = \{m_1, \dots, m_N\}$ be the constellation vector, and $\mathbf{p} = \{p_1, \dots, p_N\}$ be the power vector. Each constellation size must lie in the finite discrete set $\mathcal{M}$, i.e., $\mathbf{m} \in \mathcal{M}^N$, and each power must lie in the finite discrete set $\mathcal{P}$, i.e., $\mathbf{p} \in \mathcal{P}^N$. For example, for QAM, we set $\mathcal{M} = \{1, 2, 4, \dots, m_{\max}\}$.

Our channel model captures distance dependent path loss attenuation of signal strength:

$$h(d) = \begin{cases} d^{-\alpha}, & d > d_{\min} \\ 1, & \text{else} \end{cases}, \qquad (1)$$

where d is the distance separating the transmitting and receiving nodes, and $\alpha > 2$ is the path loss constant. The case $d < d_{\min}$ captures near-field behavior.

The channel is also subject to additive Gaussian noise which we model as a constant σ^2. The SINR seen at receiver j when listening to transmitter i at time t is

$$\mathrm{SINR}_j(t) = \frac{p_i(t) d_{ij}^{-\alpha}}{\sum_{k \in \mathcal{T}(t) \setminus \{i\}} p_j(t) d_{kj}^{-\alpha} + \sigma^2}, \qquad (2)$$

where $\mathcal{T}(t)$ is the set of nodes transmitting at time t.

The BER for receiver j when listening to transmitter i with a constellation size of $m_i(t)$ at time t is

$$\mathrm{BER}_j(t) = 2Q\Big(\sqrt{2\mathrm{SINR}(t)_j} \sin \frac{\pi}{m_i(t)}\Big). \qquad (3)$$

Finally, let there be a BER constraint $b_{\max}$ specifying the largest permissible BER on any link; $b_{\max}$ serves as a proxy for network QoS.

B. Medium access layer: globally optimal scheduling

Let time be slotted and let each round consist of a collection of S time slots $[S] = \{1, \dots, S\}$; each time slot lasts for a duration $\lambda_t, t \in [S]$. A schedule is specified by S different $N \times N$ matrices, $\mathbf{B} = (B^1, \dots, B^S)$, where

$$B_{ij}^t = \begin{cases} 1, & \text{if } i \text{ transmits to } j \text{ in slot } t \\ 0, & \text{else} \end{cases}. \qquad (4)$$

A valid schedule is one where each node i transmits in at least one time slot t.[1] Furthermore, the schedule will by construction satisfy the BER constraint for each attempted reception. That is, if $B_{ij}^t = 1$ then $\mathrm{BER}_j(t) < b_{\max}$.

Although the nominal capacity for each transmitter i is $c_i = \log_2 m_i$, each transmitter only is eligible to transmit on a total of s_i of the time slots. As such the effective capacity of link (i, j) is $\tilde{c}_{ij} = \sum_{t \in [S]} B_{ij}^t \lambda_t \log_2 m_i$.

$\mathbf{B}$ is not a free variable in the cross-layer optimization framework, and is computed prior to the resource allocation process. We employ a random packing heuristic from [2] to compute $\mathbf{B}$ such that interference is reasonably limited and every transmitter, if possible, transmits in at least one time slot.

The time-multiplexed flow graph. With the schedule in hand we form the directed directed capacitated graph $G = (V, E)$ by setting $V = \{n \in [N] : \tilde{c}_n > 0\}$ and

$$E = \{(i, j) \in V \times V \ : \ B_{ij}^t = 1 \text{ for some } t \in [S]\}, \qquad (5)$$

that is, the edge set consists of all edges carrying a transmission on one or more time slots. Finally, each edge $e = (i, j)$ is assigned a capacity $\tilde{c}_i$. Note that although G is directed, every edge (i, j) usually has an associated edge (j, i) in the reverse direction (albeit with a potentially different capacity scaling factor).

C. Network layer: multicommodity flow

Let there be K commodities, each commodity specified by a source node $\sigma_k \in [N]$, and a destination node $\delta_k \in [N]$. A *throughput vector* $\mathbf{f} = \{f_1, \dots, f_K\}$ is *feasible* if it respects both network capacity constraints and conservation of flow constraints. The objective is is to maximize the sum throughput

[1] Note that it may not be feasible for all nodes to transmit. In particular, consider a node n that is the first unassigned node for a new time slot t. If a node is unable to transmit due to the presence of excessive noise (with no interference), then the node is removed from the network as incompatible.

summed over the K commodities:

$$F(\mathbf{f}) = \sum_{k=1}^{K} f_k. \tag{6}$$

Each commodities throughput may be split across multiple paths; we define a *flow* as the set $\{x_e^k, e \in E, k \in [K]\}$, so that x_e^k is the flow for commodity k on edge e.

The centralized global optimization problem is to maximize the sum commodity flow, $F(\mathbf{f})$, subject to flow feasibility constraints and link capacity constraints. The link capacity constraints, indeed the links and nodes comprising the graph itself, depend upon the underlying temporal schedule, $\mathbf{B}$, and schedule durations λ. Finally, the feasibility of the underlying temporal schedule depends upon the physical layer power vector $\mathbf{p}$ and the constellation size vector $\mathbf{m}$. Thus, the global optimization problem is:

$$\begin{aligned}
\max_{\mathbf{p},\mathbf{m},\mathbf{x},\lambda} \quad & F(\mathbf{f}) = \sum_{k=1}^{K} f_k \\
\text{s.t.} \quad & \sum_{i\in[N]} x_{i,\delta_k}^k = f_k,\ k \in [K] \\
& \sum_{k\in[K]} x_{ij}^k \le \tilde{c}_{ij},\ i,j \in [N] \\
& \sum_{j\in[N]} x_{ji}^k = \sum_{j\in[N]} x_{ij}^k,\ k \in [K], i \in [N] \\
& \tilde{c}_{ij} = \sum_{t\in[S]} B_{ij}^t \lambda_t \log_2 m_i,\ i,j \in [N] \\
& \sum_{t\in[S]} \lambda_t = 1 \\
& B_{ij}^t \cdot \mathrm{BER}_j^t \le \beta,\ i,j \in [N] \\
& \mathrm{BER}_j^t = 2Q\Big(\sqrt{2\mathrm{SINR}_j^t}\sin\frac{\pi}{m_i}\Big),\ i,j \in [N] \\
& \mathrm{SINR}_j^t = \frac{p_i d_{ij}^{-\alpha}}{\sum_{k\in[N],k\neq i} p_k d_{kj}^{-\alpha} + \sigma^2}, i,j \in [N]
\end{aligned}$$

The first line states the objective is to maximize the sum commodity throughput. The first constraint defines the throughput for each commodity as the sum of the flow over all edges which terminate in commodity k's terminal node δ_k. The second constraint is the capacity constraint; the third is the conservation of flow constraint. The fourth constraint is the equation for transmitter effective capacity; the fifth is the requirement of schedule durations adding up to one. The sixth is the required BER constraint (only required to hold at times when the node is receiving). The seventh constraint is the expression for BER, and the last is the expression for the SINR.

The Q function, $Q(z) = \mathbb{P}(Z > z)$ for $Z \sim \mathcal{N}(0,1)$ is estimated using:

$$\begin{aligned}
z &= \sqrt{2\mathrm{SINR}_j^t}\sin\frac{\pi}{m_i}, \\
Q(z) &\approx \frac{1}{\frac{1}{2}z + \frac{1}{2}\sqrt{z^2 + \frac{8}{\pi}}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.
\end{aligned}$$

D. Three cases of design integration.

To compare network performance when optimizing in a cross-layer fashion as opposed to modularizing the framework by layer and solving each module individually until convergence, we formulate three cases where integrate and modularization is reasonable given the global optimization problem defined in the previous section:

1) MAC+NET vs. MAC|NET
2) PHY+MAC vs. PHY|MAC
3) PHY(p+m) vs. PHY($p|m$)

The comparison in all three cases is between the complete global optimization problem to a framework where that problem is split in two parts.

MAC+NET vs. MAC|NET This case splits the problem by first maximizing sum-rate link capacity $\sum_{i,j\in[N]} \tilde{c}_{ij}$ and then maximizing the flow on the resultant capacitated graph. Therefore, in this case, the split of the global framework occurs between the MAC and NET layers. The relationship between MAC and NET for this case is indicated by the | symbol.

PHY+MAC vs. PHY|MAC The optimization of power and constellation size is done independently from the optimization of the schedule time slot duration in this case, with the objective still being throughput maximization. Therefore, the split here is between the PHY and MAC layers. The two optimization problems are solved repeatedly until convergence.

PHY($\mathbf{p}+\mathbf{m}$) vs. PHY($\mathbf{p}|\mathbf{m}$) The split in this case is within the PHY layer, between power and constellation size. Each is optimized individually for maximum throughput, repeating the process until convergence.

E. Numerical solution with commercial solvers.

AMPL [13], an algebraic modeling language, was used to interface with a MINLP (mixed integer nonlinear program) solver in order to find a global optimum for the centralized network optimization problem. The data for the model was loaded from a network simulation engine developed by the authors. A specialized MINLP solver, MINLPBB [14], then parsed the AMPL-defined model and data to produce the globally optimal solution. This optimization engine uses a parallel branch-and-bound algorithm [15], [16].

III. Results

Figure 2 shows an example network topology, along with the locations of the sources and sinks of the two commodities loading the network. The graph shown is the time multiplexed flow graph obtained by the distributed scheduling algorithm. Figure 3 shows the relative improvement in sum commodity throughput as a function of the maximum acceptable BER, $b_{\max}$ for the three design integration cases. Naturally, the throughput increases as the BER constraint is relaxed. A key observation is that there is significant improvement in network performance from the solution to the integrated problem over the solution to the split dual problem.

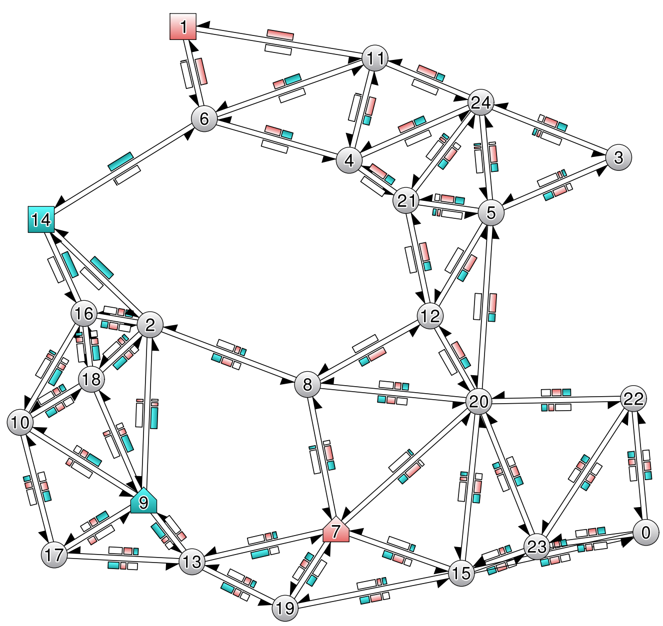

Fig. 2. A sample 25 node topology of the kind used to generate the plots. The graph is the time multiplexed flow graph obtained through the use of the distributed coloring algorithm. The network is loaded with two commodities. The house shapes (nodes 7 and 9) are the sources, the square shapes (nodes 1 and 14) are the sinks, and the circles are the relay nodes. Associated with each link is a set of boxes which designate the fraction of the capacity used by each of the commodities on that link.

IV. Conclusion

We have presented a novel formulation of a cross layer network optimization design problem, appropriate for ad hoc networks of cognitive radios. This cross layer design seeks to jointly optimize transmission power, constellation size, the temporal schedule, and the corresponding optimal flow in order to maximize the sum commodity throughput. We then demonstrated that for three variations of the modular network design, increase in network performance can be achieved by merging the two parts of the framework into a joint optimization problem.

Our future work is focused on incorporating additional components of uncertainty into our optimization engine and our distributed algorithms. In particular, we have added fading effects to our channels, random noise to the receivers, and variability in the path loss exponent.

Acknowledgments

This work is funded by the U.S. Army Communications Electronics Research, Development and Engineering Center (CERDEC), under contract #DAAB-07-01-9-L504.

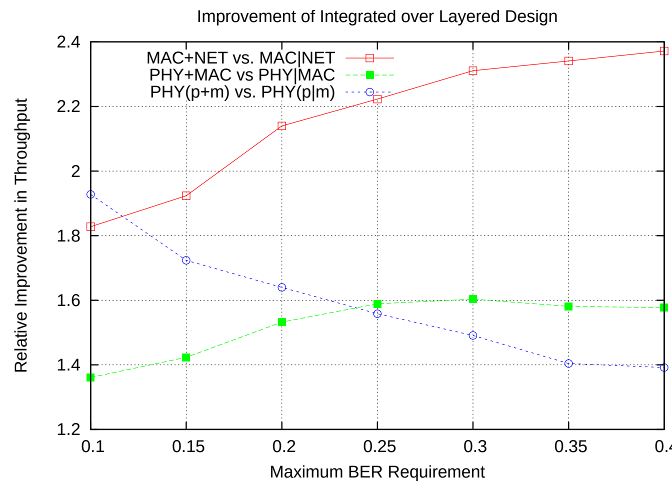

Fig. 3. Relative increase in throughput by merging the dual optimization problem into a joint one in three different cases.

References

[1] J. Mitola, "Cognitive radio: an integrated agent architecture for software defined radio," Ph.D. dissertation, Royal Institute of Technology (KTH), May 2000.

[2] Y. Wu, P. Chou, Q. Zhang, K. Jain, A. Zhu, and S.-Y. Kung, "Network planning in wireless ad hoc networks: a cross-layer approach," *IEEE JSAC*, vol. 23, no. 2, January 2005.

[3] J. Mitola and G. Maguire, "Cognitive radio: making software radios more personal," *IEEE Personal Communications Magazine*, vol. 6, no. 4, pp. 13–18, August 1999.

[4] T. Costlow, "Cognitive radios will adapt to users," *IEEE Intelligent Systems Magazine*, vol. 18, no. 3, p. 7, May–June 2003.

[5] I. Poole, "What exactly is... cognitive radio?" *IEEE Communications Engineer Magazine*, vol. 3, no. 5, pp. 42–43, October–November 2005.

[6] J. Walko, "Cognitive radio," *IEE Review Magazine*, vol. 51, no. 5, pp. 34–37, May 2005.

[7] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 1813–1827, May 2006.

[8] S. Weber, J. G. Andrews, and N. Jindal, "Spatial throughput of ad hoc networks under locally optimal sub-band selection," in *submitted to the IEEE INFOCOM*, Anchorage, AK, May 2007.

[9] ——, "Throughput and transmission capacity of ad hoc networks with channel state information," in *submitted to the 44th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, September 2006.

[10] D. P. Bertsekas, *Network optimization: continuous and discrete*. Belmont, MA: Athina Scientific, 1998.

[11] H. Luss, "Operations Research and Capacity Expansion Problems: A Survey," *Operations Research*, vol. 30, no. 5, pp. 907–947, 1982.

[12] C. YANG and J. ZHANG, "A constrained capacity expansion problem on networks," *International journal of computer mathematics*, vol. 70, no. 1, pp. 19–33, 1998.

[13] R. Fourer, D. Gay, and B. Kernighan, *AMPL: a modeling language for mathematical programming*. Thomson/Brooks/Cole, 2003.

[14] S. Leyffer, "User manual for MINLP BB," 1998.

[15] E. Lawler and D. Wood, "Branch-And-Bound Methods: A Survey," *Operations Research*, vol. 14, no. 4, pp. 699–719, 1966.

[16] B. Gendron and T. Crainic, "Parallel Branch-And-Bound Algorithms: Survey and Synthesis," *Operations Research*, vol. 42, no. 6, pp. 1042–1066, 1994.