A Distributed Optimization Algorithm for Multi-hop Cognitive Radio Networks

Yi Shi Y. Thomas Hou
The Bradley Department of Electrical and Computer Engineering
Virginia Polytechnic Institute and State University, Blacksburg, VA, 24061, USA

***Abstract*—Cognitive radio (CR) is a revolution in radio technology and is viewed as an enabling technology for dynamic spectrum access. This paper investigates how to design distributed algorithm for a future multi-hop CR network, with the objective of maximizing data rates for a set of user communication sessions. We study this problem via a cross-layer optimization approach, with joint consideration of power control, scheduling, and routing. The main contribution of this paper is the development of a distributed optimization algorithm that iteratively increases data rates for user communication sessions. During each iteration, there are two separate processes, a *Conservative Iterative Process* (CIP) and an *Aggressive Iterative Process* (AIP). For both CIP and AIP, we describe our design of routing, minimalist scheduling, and power control/scheduling modules. To evaluate the performance of the distributed optimization algorithm, we compare it to an upper bound of the objective function, since the exact optimal solution to the objective function cannot be obtained via its mixed integer nonlinear programming (MINLP) formulation. Since the achievable performance via our distributed algorithm is close to the upper bound and the optimal solution (unknown) lies between the upper bound and the feasible solution obtained by our distributed algorithm, we conclude that the results obtained by our distributed algorithm are very close to the optimal solution.**

I. Introduction

Cognitive radio (CR) is a revolution in radio technology that is enabled by recent advances in RF design, signal processing, and communications software [14]. CR is characterized by software-based signal processing and promises unprecedented flexibility in radio communications. It is considered as an enabling technology for dynamic spectrum access and its potential has been recognized by the commercial sector as well as the military (e.g., JTRS program [6]) and public safety communications (e.g., SAFECOM [15]).

Due to its software nature, a CR is capable of reconfiguring RF and switching to newly-selected frequency bands (a.k.a. frequency-agile). From wireless networking perspective, CR offers a whole new set of research problems in algorithm design and protocol implementation. To appreciate such opportunity, we compare CR with a closely related wireless technology called *multi-channel multi-radio* (MC-MR), which has also been under intensive research in recent years (see e.g., [1], [7], [8] and reference therein). First, MC-MR employs traditional *hardware-based* radio technology and thus each radio can only operate on a single channel at a time and *there is no switching of channel on a per-packet basis.* Thus, the number of concurrent channels that can be used at a wireless node is limited by the number of radios. In contrast, the radio technology in CR is software-based; a soft radio is capable of switching frequency bands on a per-packet basis. As a result, the number of concurrent frequency bands that can be used by CR is typically much larger than that can be supported by MC-MR. Second, a common assumption for MC-MR is that there is a set of "common channels" available for every node in the network; each channel typically has the same bandwidth. Such assumption is hardly true for future CR networks, in which each node may have a different set of frequency bands, each may of un-equal size. A CR node is capable of working on a set of "heterogeneous" channels that are scattered on widely-separated slices of the frequency spectrum with different bandwidths. An even more profound advance in CR technology is that future CR can work on non-contiguous channels for transmission/reception. These important differences between MC-MR and CR warrant that the algorithm design for future CR network is substantially more complex than that for current MC-MR. In some sense, an MC-MR network can be considered as a special case of a CR network. Thus, algorithms designed for future CR networks can be tailored to address current MC-MR networks while the converse is not true.

In this paper, we consider how to design distributed algorithm for future CR network. The goal is to optimize network resource utilization, with the specific objective of maximizing a common scaling factor of session's minimum rate requirement for a set of active user communication sessions. We show such problem for CR networks is inherently cross-layer in nature and calls for joint consideration of power control, scheduling, and routing. We follow the so-called "protocol model" for interference modeling [17]. Since power control directly affects the receiving power at the destination node (signal power) and at other nodes (interference power), it has profound impact on interference relationship among the nodes.

Subsequently, we move on to develop a distributed optimization algorithm, which is the main contribution of this paper. We give details of the iterative steps in the algorithm on how to increase scaling factor for user communication sessions. During each iteration, we aim to increase the smallest scaling factor among the active sessions in the network. To increase the smallest scaling factor among the sessions at each iteration, we employ two separate processes, a *Conservative Iterative Process* (CIP) and an *Aggressive Iterative Process* (AIP). Under CIP, we aim to increase the smallest scaling factor, without affecting any other session; while under AIP, we can decrease other sessions' scaling factors, as long as the affected scaling factors do not fall below the one that is being increased. The need of AIP is easy to understand. The reason why CIP is needed is interesting and will be discussed in Section III-B.

Both CIP and AIP incorporate three modules, namely routing, minimalist scheduling, and power control/scheduling. In the routing module, we define link cost based on the so-called *bandwidth-footprint product* (BFP), a unique metric associated with CR networks. In the minimalist scheduling module, scheduling assignments along the minimum cost route are made only when there is no other choices (and thus follows "minimalist" approach). The reason for this minimalist approach is that power control may change the conflict relationship among links. Therefore, scheduling assignment is best done with joint consideration of power control. Finally, the power control/scheduling module determines all the remaining scheduling assignments, transmission powers, and flow rate increase on the minimum cost route.

To evaluate the performance of the distributed optimization algorithm, we compare it to an upper bound of the objective function, since the exact optimal solution to the objective function cannot be obtained via its mixed integer nonlinear programming (MINLP) formulation. Simulation results show that the achievable performance via our distributed algorithm is close to the upper bound. Since the optimal solution (unknown) lies between the upper bound and the feasible solution obtained by our distributed algorithm, we conclude that the results obtained by our distributed algorithm are very close to the optimal solution.

The rest of this paper is organized as follows. In Section II, we present a mathematical model for power control, scheduling, and routing. In Section III, we give details on the design of a distributed optimization algorithm. In Section IV, we present simulation results for the distributed algorithm and compare the results to those from the upper bound of the objective function. Section V reviews related work and Section VI concludes this paper.

II. Problem Modeling

We consider a CR ad hoc network consisting of a set of $\mathcal{N}$ nodes. Unlike MC-MR networks where the set of available frequency bands at each node is identical, in a CR network, the set of available frequency bands at each node depends on its location and may not be the same. Denote $\mathcal{M}_i$ the set of available frequency bands at node i and $\mathcal{M}$ the union of available frequency bands among all the nodes in the network, i.e., $\mathcal{M} = \bigcup_{i\in\mathcal{N}} \mathcal{M}_i$. For each available frequency band at a node, we assume its bandwidth is W. Consider a set of $\mathcal{L}$ active user communication (unicast) sessions, each with a minimum rate requirement. Denote $s(l)$ and $d(l)$ the source and destination nodes of session $l \in \mathcal{L}$ and $r(l)$ the minimum rate requirement of session l. In this paper, we consider how to maximize these session rates proportionally. That is, we aim to find a maximum scaling factor K, such that $K \cdot r(l)$ amount of data can be transmitted for each session $l \in \mathcal{L}$.

A. Scheduling and Power Control

Scheduling for transmission at each node in the network can be done either in time domain or frequency domain. In this paper, we consider scheduling in the frequency domain in the form of frequency bands. Suppose that band m is available at both node i and node j, i.e., $m \in \mathcal{M}_i$ and $m \in \mathcal{M}_j$. Denote p_{ij}^m the transmission power from node i to node j in frequency band m. We now analyze the necessary and sufficient condition for successful transmission.

In a radio environment, there is a propagation gain associated with each transmission. Denote g_{ij} the propagation gain for transmission from node i to node j. We first consider constraint on interference. Suppose there is a transmission from node i to node j on frequency band m, then there is a limitation on the transmission power for a concurrent transmission from a neighboring node k to another node h. Specifically, we consider the received interference power on node j (due to concurrent transmission from node k to node h) is negligible only if the received interference power does not exceed a threshold P_I, i.e., $p_{kh}^m \cdot g_{kj} \le P_I$. From this, we can calculate the maximum allowed transmission power at node k (which is considered as interference at node j) as $P_{kj}^I = \frac{P_I}{g_{kj}}$. That is, to make the interference negligible on node j, we must have

$$p_{kh}^m \le P_{kj}^I \; . \tag{1}$$

With the above interference constraint, we now consider power level constraint for successful transmission. We consider that a data transmission from node i to node j on frequency band m is successful only if the received transmission power on node j exceeds a power threshold P_T, i.e., $p_{ij}^m \cdot g_{ij} \ge P_T$. From this, we can calculate the minimum required transmission power on node i as

$$p_{ij}^T = \frac{P_T}{g_{ij}} \; . \tag{2}$$

That is, to make a successful data transmission from node i to node j, we must have

$$p_{ij}^m \ge P_{ij}^T \; . \tag{3}$$

Both (1) and (3), when considered in isolation, are necessary conditions for successful transmission. But when jointly considered, they become the sufficient condition. We now generalize these conditions for successful transmission in a multi-hop network settings. To start with, we introduce the following binary indicator.

$$x_{ij}^m = \begin{cases} 1 & \text{If node } i \text{ transmits data to node } j \text{ on band } m, \\ 0 & \text{otherwise.} \end{cases}$$

As mentioned earlier, we consider scheduling in the frequency domain and thus once a band $m \in \mathcal{M}_i$ is used by node i for transmission or reception, this band cannot be used again by node i for other transmission or reception. Then we have

$$\sum_{j\in\mathcal{T}_i^m} x_{ij}^m + \sum_{k\in\mathcal{T}_i^m} x_{ki}^m \le 1 \; , \tag{4}$$

where $\mathcal{T}_i^m$ is the set of nodes to which node i can transmit under full power $P_{\max}$ on band m, i.e., $\mathcal{T}_i^m = \{j : p_{ij}^T \leq P_{\max}, j \neq i, m \in \mathcal{M}_j\}$.[1] Similarly, denote $\mathcal{I}_j^m$ the set of nodes that can produce interference on node j on band m under full power $P_{\max}$, i.e., $\mathcal{I}_j^m = \{k : P_{kj}^I \leq P_{\max}, m \in \mathcal{M}_k\}$. Based on (1) and (3), we have the following conditions for successful transmission on link $i \to j$ and interfering link $k \to h$:

$$p_{ij}^m \begin{cases} \in \left[P_{ij}^T,\ P_{\max}\right] & \text{If } x_{ij}^m = 1, \\ = 0 & \text{If } x_{ij}^m = 0. \end{cases}$$

$$p_{kh}^m \leq \begin{cases} P_{kj}^I & \text{If } x_{ij}^m = 1, \\ P_{\max} & \text{If } x_{ij}^m = 0. \end{cases} \quad (k \in \mathcal{I}_j^m, k \neq i, h \in \mathcal{T}_k^m)\ .$$

Mathematically, these conditions can be re-written as

$$p_{ij}^m \in \left[P_{ij}^T x_{ij}^m,\ P_{\max} x_{ij}^m\right] \tag{5}$$

$$p_{kh}^m \leq P_{\max} - (P_{\max} - P_{kj}^I) x_{ij}^m \quad (k \in \mathcal{I}_j^m, k \neq i, h \in \mathcal{T}_k^m)\ . \tag{6}$$

B. Flow Routing under Capacity Constraint

Recall that the objective of optimization problem is to maximize the scaling factor K so that $K \cdot r(l)$ amount of data can be transmitted for each session $l \in \mathcal{L}$. Due to limited transmission range of a node, it is necessary to employ multi-hop for data routing. Further, to achieve optimality, it is also necessary to employ *multi-path* (i.e., allow flow splitting), due to its ability for load balancing and flexibility.

Mathematically, this can be modeled as follows. Denote $f_{ij}(l)$ the data rate on link $i \to j$ that is attributed to session l, where $i \in \mathcal{N}, j \in \mathcal{T}_i = \bigcup_{m \in \mathcal{M}_i} \mathcal{T}_i^m$. If node i is the source node $s(l)$ of session l, then

$$\sum_{j \in \mathcal{T}_i} f_{ij}(l) = K \cdot r(l)\ . \tag{7}$$

If node i is the destination node $d(l)$ of session l, then

$$\sum_{k \in \mathcal{T}_i} f_{ki}(l) = K \cdot r(l)\ . \tag{8}$$

If node i is an intermediate relay node for session l, i.e., $i \neq s(l)$ and $i \neq d(l)$, then

$$\sum_{j \in \mathcal{T}_i}^{j \neq s(l)} f_{ij}(l) = \sum_{k \in \mathcal{T}_i}^{k \neq d(l)} f_{ki}(l)\ . \tag{9}$$

It can be easily verified that if (7) and (9) are satisfied, (8) must be satisfied. As a result, it is sufficient to list (7) and (9) in the formulation.

In addition to the above flow balance equations at each node $i \in \mathcal{N}$ for session $l \in \mathcal{L}$, the aggregated flow rates on each radio link cannot exceed this link's capacity, i.e., $\sum_{l \in \mathcal{L}}^{s(l) \neq j, d(l) \neq i} f_{ij}(l) \leq \sum_{m \in \mathcal{M}_{ij}} W \log_2 \left(1 + \frac{g_{ij}}{\eta W} p_{ij}^m\right)$, where η is the ambient Gaussian noise density. Note that the denominator inside the $\log$ function contains only ηW. This is due to the use of protocol interference model, i.e., when node i is transmitting to node j on band m, the interference from all other nodes in this band must remain negligible to meet scheduling constraint.

[1] We assume the peak power at each node ($P_{\max}$) is identical. Further, we assume that $g_{ij} = g_{ji}$. As a result, $\mathcal{T}_i^m$ is also the set of nodes from which node i can receive under full power $P_{\max}$ on band m.

III. Design of A Distributed Optimization Algorithm

In this section, we present a distributed optimization algorithm for our problem. This algorithm increases the scaling factor K iteratively and terminates when K cannot be further increased. The main idea of this algorithm is given in Section III-A, which includes session selection, routing, minimalist scheduling, and power control/scheduling modules. The details of each module are described in Section III-B.

A. Main Idea

Our distributed algorithm increases the scaling factor K iteratively and terminates when K cannot be further increased. In a distributed network, each session $l \in \mathcal{L}$ maintains its own current scaling factor $K(l)$, which may be different from other sessions. During each iteration, we find the smallest scaling factor among all sessions in $\mathcal{L}$ and attempt to increase it. When there are multiple sessions with the same smallest scaling factor, we break the tie deterministically based on their source node IDs. The benefit of the deterministic approach is that a unique session is chosen in our distributed algorithm. For example, we could choose the session with the smallest source node ID among tied sessions. For implementation, this session selection step can be done via broadcast by the source node of each session $l \in \mathcal{L}$.

Upon identifying the session l with the smallest current scaling factor $K(l)$, we move onto iteration process shown in Fig. 1. There are two separate processes, namely, a *Conservative Iterative Process* (CIP) on the left and an *Aggressive Iterative Process* (AIP) on the right. Although both processes contain routing, minimalist scheduling, and power control/scheduling modules, they differ in both objectives as well as details. The objective of CIP is to increase $K(l)$ without affecting (decreasing) the current scaling factors of other sessions in $\mathcal{L}$. On the other hand, the objective of AIP is to increase $K(l)$ aggressively by decreasing the current scaling factors of some other sessions in $\mathcal{L}$ as long as they do not fall below the newly increased $K(l)$. The need of AIP is easy to understand. So the question becomes why CIP is necessary. The answer to this question lies in the way we define link cost in AIP-Routing module, which is different from that in CIP-Routing module. We will explain in detail how the definition of link cost in AIP-Routing module mandates the need of CIP in Section III-B (Remark 1).

We first present the ideas in the routing module under CIP and AIP. During an iteration, the routing, scheduling, and power control for session l in the previous iterations are intact. The CIP-Routing module aims to find an additional route (which could overlap with previous routes) for session l onto which there is a potential to push more data rate. This routing module is based on minimum cost, which can be implemented distributedly. The key step in CIP-Routing is the definition of incremental link cost (ILC) for pushing more data rate

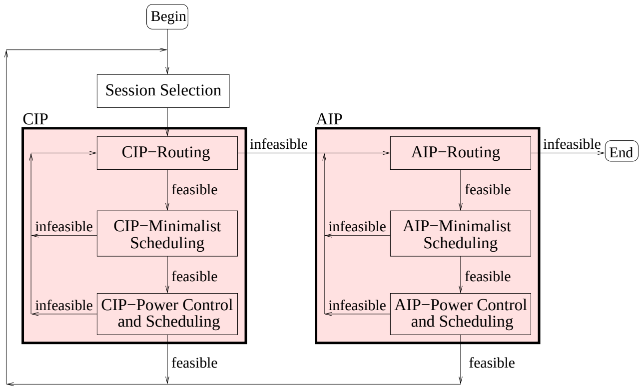

Fig. 1. Flow chart of our algorithm.

onto a link. Obviously, such link cost must capture network resource in terms of both frequency usage (bandwidth) and spatial occupancy (footprint). In light of this spirit, we use the so-called *bandwidth-footprint product* (BFP) [17], which is a unique metric for CR networks. So the incremental link cost for pushing more data rate onto a link can be defined as the incremental BFP per additional data rate. This metric only requires local information and can be computed distributedly. On the other hand, under the AIP-Routing module, all links carrying sessions whose current scaling factors are greater than $K(l)$ will be marked. The cost on these links will be redefined so that session l has the potential of pushing more data rate at the expense of decreasing the data rate of those sessions currently with larger scaling factors.

We now present the ideas in the minimalist scheduling module under CIP and AIP. Our approach is of "minimalist," in the sense that we only make necessary scheduling decisions (i.e., frequency band assignments) when there is no alternative options. Specifically, under the CIP-Minimalist Scheduling, if there is no remaining capacity on a hop and current transmission powers on used bands have already reached their maximum allowed transmission power, then it is necessary to assign a new band. If there is only one unassigned band on this link, we will make an assignment of this band (as there is no other options) subject to scheduling constraint at the node. On the other hand, when there are multiple unassigned bands, the minimalist approach calls for deferring band assignment in the power control/scheduling module (to be discussed in the next paragraph). The reason for this deferring is that power control may change the conflict relationship among the nodes. Therefore, scheduling decision (band assignment among multiple unassigned bands) is best done with joint consideration of power control. The AIP-Minimalist Scheduling module follows a similar process, with the difference being when a new band should be assigned. This is because under AIP-Minimalist Scheduling, if a hop carries sessions with their current scaling factors greater than $K(l)$, then there is no need to assign a new band since the rates of these sessions can be reduced and thus leave more room for increasing the rate of session l.

Finally, we discuss the power control/scheduling module under CIP and AIP. This module sets the transmission power on a currently active band or some new band along the minimum cost route chosen in the routing module. The objective is to allow some additional flow rate $f(l)$ to be transmitted on this route for session l. The specific value of $f(l)$ can be determined hop by hop along the minimum cost route. Under CIP-Power Control/Scheduling, at each node along the route, it tries the following strategies to accommodate $f(l)$: (i) use the remaining capacity on this hop if possible; (ii) increase power on a currently active band so as to increase link capacity; and (iii) activate a new frequency band (previously unassigned). Band selection in the last strategy is one that is missing in the minimalist scheduling approach discussed in the last paragraph for new band assignment (when there are multiple unassigned bands). It is also important to realize that when a new band is assigned, the maximum allowed transmission power at a nearby node on this band may need to be reduced by (6). The AIP-Power Control/Scheduling module is similar to CIP-Power Control/Scheduling module, but with one more strategy to accommodate $f(l)$. That is, after (i) and before (ii), AIP-Power Control/Scheduling will check whether $f(l)$ can be accommodated by the newly released capacity from sessions currently with larger scaling factors.

It should be clear that the smallest scaling factor among the active sessions will be increased after either CIP or AIP is completed successfully after an iteration.

B. Details of Each Module

Before we present the details in our algorithm, we first introduce the following notation. For the distributed algorithm, we re-define x_{ij}^m as follows.

$$x_{ij}^m = \begin{cases} 1 & \text{If band } m \text{ is used (assigned) on link } i \to j. \\ 0 & \text{If band } m \text{ is unassigned on link } i \to j. \\ -1 & \text{If band } m \text{ cannot be used on link } i \to j. \end{cases}$$

Another notation we need in the iteration of the distributed algorithm is the maximum allowed transmission power $(p_{ij}^m)_U$. Recall that $P_{\max}$ is the maximum transmission power at a node. During an iteration, we may find that, under various constraints, the current maximum allowed transmission power may be smaller than $P_{\max}$. We use $(p_{ij}^m)_U$ for this purpose, where subscript U indicates the current upper bound on the transmission power.

Routing Module. As discussed in Section III-A, the key step in the routing module is the definition of link cost that captures network resource usage. We use bandwidth-footprint product (BFP) for this purpose, which is a unique metric for CR networks [17]. For our problem, since each band has the same bandwidth, BFP reduces to footprint, which is the interference area for a transmission. Therefore, the definition of *incremental link cost* (ILC) becomes the additional required footprint over the increase of flow rate for the session with the smallest scaling factor.

To compute footprint area, we use the widely-used propagation gain model $g_{ij} = d_{ij}^{-\alpha}$, where d_{ij} is the physical distance between nodes i and j and α is the path loss index. For a transmission from node i to node j on band m, based on (6), a node h is interfered when $p_{ij}^m > P_{ih}^I = \frac{P_I}{g_{ih}}$. Thus, the interference range of node i on band m is $\left(\frac{p_{ij}^m}{P_I}\right)^{1/\alpha}$, which is band-dependent.

Since a wireless link in a CR network is associated with multiple frequency bands, the computation of ILC must be somehow related to the cost of each frequency band. We give some details here on how ILC is computed. Initially, for band m on link $i \to j$, since $p_{ij}^m = 0$ (zero transmission power), both the capacity on this band and the footprint area are 0. We define the *incremental band cost* (IBC) as the incremental footprint area over increased band capacity when the transmission power is increased to the minimum required transmission power P_{ij}^T (see (2)). That is, the IBC for band m on link $i \to j$ is

$$IBC(i,j,m) = \frac{\pi(P_{ij}^T/P_I)^{2/\alpha}}{W\log_2(1+\frac{g_{ij}P_{ij}^T}{\eta W})} \, . \tag{10}$$

Since the above IBC is identical on all bands initially, we can define ILC as the IBC on any band.

In subsequent iterations, the definition of ILC is case specific. For the simple case where link $i \to j$ has a positive remaining capacity (the remaining capacity c_{ij} is the sum of capacity on each band minus the total session flow rates on this link), the ILC is defined as 0, since this link can support additional flow rate without increasing its transmission power (or its footprint). For other cases (i.e., no remaining capacity), the computation of IBC depends on whether or not the band is currently used.

- Case I: If band m is already used but $p_{ij}^m < (p_{ij}^m)_U$, then p_{ij}^m may be increased to $(p_{ij}^m)_U$ and IBC is defined as

$$IBC(i,j,m) = \frac{\pi[(p_{ij}^m)_U/P_I]^{2/\alpha} - \pi(p_{ij}^m/P_I)^{2/\alpha}}{W[\log_2[1+\frac{g_{ij}}{\eta W}(p_{ij}^m)_U] - \log_2(1+\frac{g_{ij}}{\eta W}p_{ij}^m)]} \, . \tag{11}$$

- Case II: If band m is not yet used, then IBC can be defined by (10).
- Case III: If band m is already fully utilized (i.e., $(p_{ij}^m)_U = p_{ij}^m$), IBC is then defined as ∞, since the capacity on this band cannot be further increased.

Since on link $i \to j$, different band may have different IBC, we need to have a band selection policy to decide which band to use and subsequently to define ILC based on the chosen band. The key idea on our band selection policy is to use a band that has already been used to its fullest extent before considering deploying a new band. Under this spirit, when there exists a Case I band, we will choose such a band with the smallest IBC. As a result, ILC will be defined the same as the IBC of such band. Otherwise, we examine if there exist a Case II band and will use it if available. As a result, ILC will be defined as the IBC of such a Case II band. When neither Case I nor Case II band exists, ILC will be defined as ∞, since this link's capacity cannot be further increased. Clearly, ILC can be computed locally, since (10) and (11) can be computed locally at node i. The pseudo-code for ILC computation in CIP-Routing is given in [18].

For the AIP-Routing module, we have the additional flexibility of reducing flow rates of some sessions with larger scaling factors so as to increase the smallest scaling factor of a session under consideration. As a result, the ILC definition needs to consider such scenario. For a session with a larger scaling factor, we need to make sure that, after a reduction, its scaling factor will no fall below the scaling factor of the session that has been increased. Under this constraint, we can compute the capacity that can be released. In the case when there exist multiple sessions with larger scaling factors, we will choose the session with the largest releasable capacity. Note that in this scenario, there is no need to increase the transmission power (and footprint) of any band on this link. Thus, ILC will be 0 since there is no change in footprint.

Under both CIP-Routing and AIP-Routing, ILC of a link may be different during each iteration. As a result, the minimum cost route at each iteration could be different for the same session. The union of these routes for all the iterations will lead to a multi-path routing solution for a session, which is important in terms of maximizing our objective K.

Remark 1: With link cost definitions in CIP and AIP, we can now explain our earlier question of why CIP is needed in our algorithm (see Fig. 1). Note that under AIP, any link that has sessions with larger scaling factors than the session under optimization will have a 0 *link cost. If AIP is used alone, then many sessions may attempt to traverse such* 0*-cost link, making such links bottleneck in the network. By using CIP before AIP, we are able to distribute sessions more evenly among the network without getting into such bottleneck situation. Therefore, CIP is an essential mechanism for the proper operation of the distributed optimization algorithm while AIP is only used as an enhancement mechanism.* □

Minimalist Scheduling Module. As described in Section III-A, we follow a minimalist approach in scheduling (band assignments), i.e., a band is assigned only when there is no other alternatives.

An important observation is that when we perform scheduling, only those nodes that are within one hop away (under $P_{\max}$) along the minimum cost route may be affected. This is because that once we assign a new band m on a link $i \to j$, this band can no longer be used again by either node i or node j for transmission/reception. Further, for nodes that are one-hop neighbors of node i or node j, band m can no longer be used for transmission to node i or node j. Thus, the effect of scheduling assignment at node i on band m is limited to those nodes that are within one hop from either node i or node j. Note that the fact that band m is a new band (previously unassigned), i.e., available for assignment on link $i \to j$, infers that band m must be an unassigned band on the neighboring nodes. Thus, it is always feasible to assign band m unusable at neighboring nodes. By applying the same scheduling approach to all the nodes along the route, it is not hard to see that only those nodes that are within one hop away from the nodes along the minimum cost route may be affected.

We now describe how the minimalist scheduling is done along the minimum cost route from source to destination. In the CIP-Minimalist Scheduling module, for each hop $i \to j$, if node i finds that there is no remaining capacity on link $i \to j$ and transmission power on all used bands cannot be increased, then it is necessary to assign a new band on this link. In this case, if there is only one remaining band for assignment, then this band must be assigned now, i.e., $x_{ij}^m = 1$.

Upon the assignment of a new band on link $i \to j$, it is necessary to make updates in both backward (toward the source) and forward (toward the destination) directions. In the backward direction, based on scheduling constraint, this band must not be used on the last hop, say $k \to i$. If band m is an unassigned band on this link, then node k assigns band m to be unusable, i.e., $x_{ki}^m = -1$. Due to this operation (removal of band m on link $k \to i$), we may be in a situation where we can now make a band assignment on link $k \to i$. If so, we will make this assignment. Subsequently after such new band assignment along backward route, it is necessary to further go backward to make new updates.

In the forward direction, the update follows the same token, except that we may encounter an infeasible situation. In this case, link $i \to j$ should be removed from future minimum cost routing by setting its ILC to ∞. Further, all the scheduling assignments done previously from the source node to the current node will be revoked. The pseudo-code for CIP-Minimalist Scheduling is given in [18].

For the AIP-Minimalist Scheduling module, since we have the additional flexibility of reducing the flow rate of some other sessions with larger scaling factors, the decision of when a new band must be assigned will thus differ from that in the CIP-Minimalist Scheduling module. Recall that in CIP-Minimalist Scheduling, a new band should be assigned if there is no remaining capacity on a link and transmission power on all used bands cannot be increased. In contrast, in AIP-Minimalist Scheduling, we will consider the assignment of a new band only when the link does not have any other sessions with larger scaling factors, in addition to those conditions in CIP-Minimalist Scheduling.

Power Control/Scheduling Module. The last module in either CIP or AIP is power control/scheduling. In this module, we will determine all the remaining scheduling assignments (that are not determined in the minimalist scheduling module), transmission powers, and flow rate increase on the minimum cost route.

Again, power control/scheduling is performed on a hop by hop basis along the route from source to destination. The potential increase in session flow rate $f(l)$ can also be computed hop by hop. Under CIP-Power Control/Scheduling, this is accomplished with the following steps.

- Step 1 For link $i \to j$, if there is positive remaining capacity c_{ij} on this link, then node i updates the flow rate $f(l)$ with $\min\{c_{ij}, f(l)\}$.
- Step 2 Else, if a band m is assigned in the minimalist scheduling module, then node i will first check whether or not this band remains available (after power control operation along the route).
  - (2A) If this band is still available, then node i uses the transmission power P_{ij}^T. Node i also needs to update either flow rate or the remaining capacity on this link.
  - (2B) If this band is no longer available, then we come to an infeasible situation. In this case, link $i \to j$ should be removed from future minimum cost routing. This can be done by setting its ILC to ∞. Further, all the power control/scheduling done previously from the source node to the current node will be revoked.
- Step 3 Else, if it is possible to increase the transmission power on bands that are already used, then node i chooses a band m among such bands with the smallest IBC. On band m, node i will increase its transmission power to a level that can support current flow rate $f(l)$ or the maximum allowed transmission power $(p_{ij}^m)_U$, whichever is smaller.
- Step 4 Else, node i will use a new band with the largest maximum allowed transmission power under the condition that such band is still available for assignment at this point. This band assignment fills the gap in the previous minimalist scheduling module. Subsequently, we perform the same tasks in Step 2A (if an available band is found) or 2B (if no available band).

The pseudo-code for CIP-Power Control/Scheduling is given in [18]. We emphasize the benefit of keep tracking of the maximum allowed transmission power at each node. As long as power level does not exceed this limit, each node is guaranteed not to cause non-negligible interference at any other active receiver.

For the AIP-Power Control/Scheduling module, we have one more strategy to explore in order to accommodate the additional flow rate $f(l)$. That is, after Step 1 and before Step 2, if there are other sessions with larger scaling factors on this link, then we can obtain some additional capacity by reducing the scaling factor of one of these sessions. Among these sessions, we choose the one with the largest releasable capacity. For this session, we also need to reduce its flow rate on other links along its paths. The transmission power and scheduling on these links may also need to be updated. The details are given in [18].

IV. Simulation Results

In this section, we present simulation results to demonstrate the performance of our distributed algorithm. Ideally, the best performance measure of our distributed optimization algorithm is the solution to the centralized optimization problem. However, as shown in the Appendix section, the problem formulation is in the form of mixed integer nonlinear program (MINLP) and is NP-hard. Although the exact solution to the

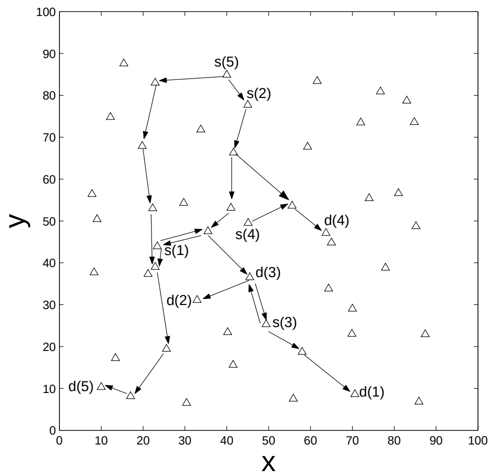

Fig. 2. 50-node CR network with 5 active sessions.

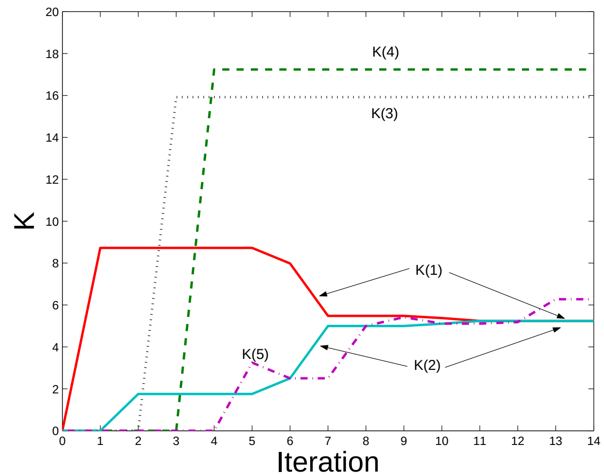

Fig. 3. Iterations for scaling factor of each session during the simulation.

MINLP problem is not obtainable, we are able to develop a tight upper bound for the objective function of the problem via linear relaxation. As a result, we will compare the performance of our distributed optimization algorithm to this centralized upper bound. Simulation results show that the achievable performance via our distributed algorithm is close to the upper bound. Since the optimal solution (unknown) lies between the upper bound and the feasible solution obtained by our distributed algorithm, we conclude that the results obtained by our distributed algorithm are very close to the optimal solution.

A. Simulation Setting

We consider $|\mathcal{N}| = 20, 30, 40,$ and 50 nodes randomly deployed in a 100x100 area. The units for distance, rate, and power density are all normalized with appropriate dimensions. Among these nodes, there are $|\mathcal{L}| = 3$ or 5 active sessions, with source node and destination node of each session randomly selected and the minimum rate requirement is 10.

There are $|\mathcal{M}| = 10$ different frequency bands in the network. However, at each node, only a subset of these frequency bands may be available. In the simulation, this is done by randomly selecting a subset of bands for each node from the pool of 10 bands. Each band has a bandwidth of $W = 50$.

We assume that the maximum transmission range on each node is 20. Correspondingly, the maximum transmission power is then $P_{\max} = (20)^\alpha P_T$, where the path loss index α is taken to be 4. Let the transmission threshold $P_T = \eta W = 50\eta$. We assume the interference range is twice of the transmission range. Then the interference threshold is $P_I = \left(\frac{1}{2}\right)^\alpha P_T = \frac{50}{16}\eta$.

B. A Case Study

Before we present complete simulation results, we first examine the iterative behavior and convergence of our distributed algorithm via a simulation case study. For this purpose, we consider the case of 50-node network shown in Fig. 2 with 5 sessions (source and destination of each session are also marked in the figure, e.g., $s(1)$ and $d(1)$ denote source and destination nodes of session 1, respectively).

Initially, all sessions start with their scaling factors as 0. Thus, the session with the smallest source node ID is chosen, say session 1. A minimum cost route, as well as scheduling and power control, for this session is constructed and its scaling factor is increased to $K(1) = 8.73$ (corresponding to session rate of 87.3). At the second iteration, there are 4 sessions with scaling factors 0. Among then, session 2 is chosen and its scaling factor $K(2)$ is increased to 1.75. After five iterations, we have $K(1) = 8.73, K(2) = 1.75, K(3) = 15.92, K(4) = 17.24$, and $K(5) = 3.25$. Since $K(2) = 1.75$ is the smallest among the five, we choose session 2 and try to increase its scaling factor in the next (sixth) iteration. During the sixth iteration, we find that it is infeasible to increase $K(2)$ under CIP. Thus, we resort to AIP, i.e., try to release some capacity from other sessions such that $K(2)$ may be increased. We find that $K(2)$ can be increased to 2.50 by decreasing $K(1)$ to 7.98 and $K(5)$ to 2.50. This iteration process continues. Finally, at the 14-th iteration, we find that $K(2) = 5.24$ is the smallest scaling factor among the sessions. However, we find that it is not feasible to increase its scaling factor under either CIP or AIP. Thus, our algorithm terminates at this iteration.

Note that our distributed algorithm offers a multi-path routing solution as we designed. In this network, session 5 uses two paths for flow routing in the final solution (see Fig. 2).

We now examine the performance of the distributed iteration. Using the centralized upper bound calculation in the Appendix, we find that the upper bound for the scaling factor is 5.42. In our distributed algorithm, we have achieved a scaling factor of 5.24, which is 96.7% of the upper bound. Although the maximum achievable (optimal) scaling factor is unknown, it is no more than this upper bound (5.42). Therefore, we conclude that the scaling factor achieved by our distributed algorithm is *at least* 96.7% of the maximum.

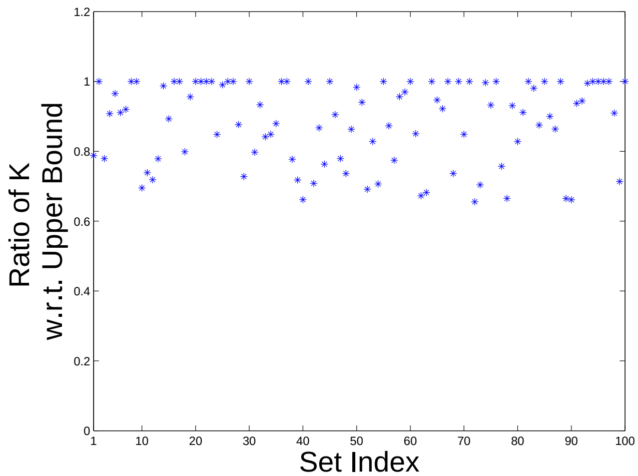

Fig. 4. Ratio of K w.r.t. upper bound for 100 data sets.

C. Complete Simulation Results

We now show complete simulation results for all 100 simulation runs. Figure 4 shows the ratio of scaling factor K obtained via our distributed algorithm over that obtained by the upper bound. The average ratio for these 100 data is 88.6%, with a standard derivation of 11.5%. We find that half of these ratio are at least 91.1%. Since the maximum achievable scaling factor (unknown) lies between the upper bound and the feasible solution obtained by our distributed algorithm, we can conclude that (i) the upper bound is very tight; and (ii) the scaling factor obtained by our distributed algorithm must be even closer to the optimal solution.

V. Related Work

Related work on MC-MR network and its relationship to CR network have been discussed in Section I. In this section, our review focus on two lines of research that are relevant to this research: (1) recent advances in multi-hop CR networks; and (2) distributed optimization algorithms for wireless networks.

For multi-hop networking with CRs, there is limited amount of work to date in the literature. In [22], Zhao et al. designed a distributed spectrum sharing approach and showed that this approach offers throughput improvement over a dedicated channel approach. In [20], Ugarte and McDonald found an upper bound on network capacity for multi-hop CR networks, although it is not clear how tight this bound is. In [21], Xin et al. studied how to assign frequency bands at each node to form a topology such that a certain performance metric can be optimized. The authors considered the so-called fixed channel approach whereby the radio is assumed to operate on only one channel at a specific time. In [19], Steenstrup studied three different frequency assignment problems. Each is viewed as a graph-coloring problem, and both centralized and distributed algorithms are presented. In [17], Shi and Hou studied the impact of power control on scheduling, routing, and network resource allocation for a CR network via a centralized approach.

There has been active research on distributed optimization algorithms for wireless networks. Some of these algorithms focus on routing problem (e.g., [11]) or scheduling problem (e.g., [2], [3], [13]), without consideration of cross-layer optimization. Cross-layer optimization problems considering joint routing and scheduling include those in [4], [10]. In these efforts, however, power control is not part of the optimization space. With fixed power, both the interference relationship among links and the capacity of each link can be computed easily. In [9], [12], power control has been considered as part of the cross-layer problems. Specifically, in [9], Lin and Shroff designed a distributed algorithm for maximizing total session utility under fixed route for each session. Their algorithm was shown to achieve a constant factor of the capacity region. But routing was not part of the optimization space. In [12], Palomar and Chiang solved some maximizing network utility problems via some new distributed decomposition approaches. In these problems, scheduling was not considered and the only power constraint is that the total transmission power at all links is bounded. As expected, joint consideration of power control, scheduling, and routing will make underlying distributed optimization problem much more complex.

VI. Conclusion

In this paper, we investigated how to design distributed optimization algorithm for a future multi-hop cognitive radio network. We first developed a mathematical model for such problem with joint consideration of power control, scheduling, and routing. The main contribution of this paper is the development of a distributed optimization algorithm that iteratively increases the scaling factor of a session. This algorithm consists of a conservative iterative process and an aggressive iterative process, each of which incorporates modules for routing, minimalist scheduling, and power control/scheduling. Through simulation results, we compared the performance of the distributed optimization algorithm with an upper bound and validated its efficacy.

Appendix – A Centralized Problem Formulation and Upper Bound

To develop a performance benchmark for the distributed optimization algorithm, we first formulate the cross-layer optimization problem. Putting together all the constraints for power control, scheduling, and routing described in Section II, we have the following formulation.

$$
\begin{array}{lll}
\text{Max} & K & \\
\text{s.t.} & \displaystyle\sum_{j \in \mathcal{T}_i^m} x_{ij}^m + \sum_{k \in \mathcal{T}_i^m} x_{ki}^m \le 1 & (i \in \mathcal{N}, m \in \mathcal{M}_i) \\
& p_{ij}^m - P_{ij}^T x_{ij}^m \ge 0 & (i \in \mathcal{N}, m \in \mathcal{M}_i, j \in \mathcal{T}_i^m) \quad (12) \\
& p_{ij}^m - P_{\max} x_{ij}^m \le 0 & (i \in \mathcal{N}, m \in \mathcal{M}_i, j \in \mathcal{T}_i^m) \quad (13)
\end{array}
$$

$$\sum_{h\in\mathcal{T}_k^m} p_{kh}^m + (P_{\max} - P_{kj}^I)x_{ij}^m \le P_{\max}$$
$$(i\in\mathcal{N}, m\in\mathcal{M}_i, j\in\mathcal{T}_i^m, k\in\mathcal{I}_j^m, k\neq i) \quad (14)$$

$$\sum_{l\in\mathcal{L}}^{s(l)\neq j, d(l)\neq i} f_{ij}(l) - \sum_{m\in\mathcal{M}_{ij}} W\log_2\left(1+\frac{g_{ij}}{\eta W}p_{ij}^m\right) \le 0$$
$$(i\in\mathcal{N}, j\in\mathcal{T}_i) \quad (15)$$

$$\sum_{j\in\mathcal{T}_i} f_{ij}(l) - K\cdot r(l) = 0 \qquad (l\in\mathcal{L}, i = s(l))$$
$$\sum_{j\in\mathcal{T}_i}^{j\neq s(l)} f_{ij}(l) - \sum_{k\in\mathcal{T}_i}^{k\neq d(l)} f_{ki}(l) = 0 \quad (l\in\mathcal{L}, i\in\mathcal{N}, i\neq s(l), d(l))$$
$$x_{ij}^m \in \{0,1\}, p_{ij}^m \ge 0 \quad (i\in\mathcal{N}, m\in\mathcal{M}_i, j\in\mathcal{T}_i^m)$$
$$f_{ij}(l), K \ge 0 \quad (l\in\mathcal{L}, i\in\mathcal{N}, i\neq d(l), j\in\mathcal{T}_i, j\neq s(l)),$$

where $P_{ij}^T, P_{kj}^I, P_{\max}, W, g_{ij}, \eta$, and $r(l)$ are all constants and $x_{ij}^m, p_{ij}^m, f_{ij}(l)$, and K are all optimization variables. In this formulation, (12) and (13) come from (5). Constraint (14) is based on (6) by noting that in (4), there is at most one $x_{kh}^m = 1$ for $h \in \mathcal{T}_k^m$. As a result, based on (13), there is at most one $p_{kh}^m > 0$ for $h \in \mathcal{T}_k^m$. Thus, (6) can be rewritten as $\sum_{h\in\mathcal{T}_k^m} p_{kh}^m \le P_{\max} - (P_{\max} - P_{kj}^I)x_{ij}^m$ for $k\in\mathcal{I}_j^m$ and $k\neq i$, which is equivalent to (14).

This optimization problem is in the form of *mixed-integer non-linear program* (MINLP), which is NP-hard in general [5]. Since we are not able to solve it exactly, we develop an upper bound to this problem, which can be used as a measure of the performance of the distributed optimization algorithm.

An upper bound for the MINLP problem can be found via a linear relaxation. This process involves two components. First, for integer variables x_{ij}^m, we can simply relax them as continuous variables defined over $[0,1]$. Second, for $\log$ term in (15), we propose to employ three tangential supports as its approximation, which is a convex envelope linear relaxation. That is, for each p_{ij}^m, we introduce a new variable v_{ij}^m and let $v_{ij}^m = \log_2\left(1+\frac{g_{ij}}{\eta W}p_{ij}^m\right)$. A linear relaxation can be obtained by replacing $\log_2\left(1+\frac{g_{ij}}{\eta W}p_{ij}^m\right)$ by v_{ij}^m and adding the following linear constraints on p_{ij}^m and v_{ij}^m [16]

$$v_{ij}^m - \tfrac{g_{ij}}{\eta W\ln 2}p_{ij}^m \le 0$$
$$v_{ij}^m - \tfrac{g_{ij}}{(\eta W + g_{ij}\beta)\ln 2}p_{ij}^m \le \log_2\left(1+\tfrac{g_{ij}\beta}{\eta W}\right) - \tfrac{g_{ij}\beta}{(\eta W+g_{ij}\beta)\ln 2}$$
$$v_{ij}^m - \tfrac{g_{ij}}{(\eta W + g_{ij}P_{\max})\ln 2}p_{ij}^m \le \log_2\left(1+\tfrac{g_{ij}P_{\max}}{\eta W}\right) - \tfrac{g_{ij}P_{\max}}{(\eta W+g_{ij}P_{\max})\ln 2}$$
$$v_{ij}^m - \tfrac{\log_2[1+g_{ij}P_{\max}/(\eta W)]}{P_{\max}}p_{ij}^m \ge 0$$

where $\beta = \left[\log_2\left(1+\frac{g_{ij}P_{\max}}{\eta W}\right) - \frac{g_{ij}P_{\max}}{(\eta W+g_{ij}P_{\max})\ln 2}\right] / \left[\frac{g_{ij}}{\eta W\ln 2} - \frac{g_{ij}}{(\eta W+g_{ij}P_{\max})\ln 2}\right]$. As a result, the non-polynomial ($\log$) term can be relaxed into linear constraints.

Based on the above process, we have a linear relaxation and its solution can be obtained by solving a single linear program.

ACKNOWLEDGMENT

This work was supported in part by NSF Grants CNS-0721570 and CNS-0627436.

REFERENCES

[1] M. Alicherry, R. Bhatia, and L. Li, "Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks," in *Proc. ACM Mobicom,* pp. 58–72, Cologne, Germany, Aug. 28–Sep. 2, 2005.

[2] A. Brzezinski, G. Zussman, and E. Modiano, "Enabling distributed throughput maximization in wireless mesh networks – A partitioning approach," in *Proc. ACM Mobicom,* pp. 26–37, Los Angeles, CA, Sep. 24–29, 2006.

[3] M. Chiang, S. Zhang, and P. Hande, "Distributed rate allocation for inelastic flows: Optimization frameworks, optimality conditions, and optimal algorithms," in *Proc. IEEE Infocom,* pp. 2679–2690, Miami, FL, March 13–17, 2005.

[4] A. Eryilmaz, A. Ozdaglar, and E. Modiano, "Polynomial complexity algorithms for full utilization of multi-hop wireless networks," in *Proc. IEEE Infocom,* Anchorage, AL, May 6–12, 2007.

[5] M.R. Garey and D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-completeness, W.H. Freeman and Company,* pp. 245–248, New York, NY, 1979.

[6] Joint Tactical Radio System, http://enterprise.spawar.navy.mil/body.cfm?type=c&category=27&subcat=60

[7] M. Kodialam and T. Nandagopal, "Characterizing the capacity region in multi-radio multi-channel wireless mesh networks," in *Proc. ACM Mobicom,* pp. 73–87, Cologne, Germany, Aug. 28–Sep. 2, 2005.

[8] P. Kyasanur and N.H. Vaidya, "Capacity of multi-channel wireless networks: impact of number of channels and interfaces," in *Proc. ACM Mobicom,* pp. 43–57, Cologne, Germany, Aug. 28–Sep. 2, 2005.

[9] X. Lin and N.B. Shroff, "The impact of imperfect scheduling on cross-layer rate control in wireless networks," in *Proc. IEEE Infocom,* pp. 1804–1814, Miami, FL, March 13–17, 2005.

[10] X. Lin and S. Rasool, "A distributed joint channel-assignment, scheduling and routing algorithm for multi-channel ad hoc wireless networks," in *Proc. IEEE Infocom,* Anchorage, AL, May 6–12, 2007.

[11] Z. Li and B. Li, "Efficient and distributed computation of maximum multicast rates," in *Proc. IEEE Infocom,* pp. 1618–1628, Miami, FL, March 13–17, 2005.

[12] D.P. Palomar and M. Chiang, "Alternative decompositions for distributed maximization of network utility: framework and applications," in *Proc. IEEE Infocom,* Barcelona, Catalunya, Spain, April 23–29, 2006.

[13] C. Peng, H. Zheng, and B.Y. Zhao, "Utilization and fairness in spectrum assignment for opportunistic spectrum access," *ACM/Springer MONET,* vol. 11, issue 4, pp. 555–576, Aug. 2006.

[14] J.H. Reed, *Software Radio: A Modern Approach to Radio Engineering,* Prentice Hall, May 2002.

[15] SAFECOM — http://www.safecomprogram.gov/SAFECOM/.

[16] Y. Shi, Y.T. Hou, H.D. Sherali, and S. Kompella, "Cross-layer optimization for UWB-based ad hoc networks," in *Proc. IEEE Military Communications Conference,* pp. 1–7, Washington, DC, Oct. 23–25, 2006.

[17] Y. Shi and Y.T. Hou, "Optimal power control for multi-hop software defined radio networks," in *Proc. IEEE Infocom,* Anchorage, AL, May 6–12, 2007.

[18] Y. Shi and Y.T. Hou, "A distributed optimization algorithm for multi-hop cognitive radio networks," Technical Report, the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, June 2007.

[19] M.E. Steenstrup, "Opportunistic use of radio-frequency spectrum: A network perspective," in *Proc. IEEE DySpan,* pp. 638–641, Baltimore, MD, Nov. 8–11, 2005.

[20] D. Ugarte and A.B. McDonald, "On the capacity of dynamic spectrum access enabled networks," in *Proc. IEEE DySpan,* pp. 630–633, Baltimore, MD, Nov. 8–11, 2005.

[21] C. Xin, B. Xie, and C.-C. Shen, "A novel layered graph model for topology formation and routing in dynamic spectrum access networks," in *Proc. IEEE DySpan,* pp. 308–317, Baltimore, MD, Nov. 8–11, 2005.

[22] J. Zhao, H. Zheng, and G. Yang, "Distributed coordination in dynamic spectrum allocation networks," in *Proc. IEEE DySpan,* pp. 259–268, Baltimore, MD, Nov. 8–11, 2005.