Beyond Overlay: Reaping Mutual Benefits for Primary and Secondary Networks Through Node-Level Cooperation

Xu Yuan, *Member, IEEE*, Yi Shi, *Senior Member, IEEE*, Xiaoqi Qin, *Student Member, IEEE*, Y. Thomas Hou, *Fellow, IEEE*, Wenjing Lou, *Fellow, IEEE*, Sastry Kompella, *Senior Member, IEEE*, Scott F. Midkiff, *Senior Member, IEEE*, and Jeffrey H. Reed, *Fellow, IEEE*

Abstract—Existing spectrum sharing paradigms have set clear boundaries between the primary and secondary networks. There is either no or very limited node-level cooperation between the primary and secondary networks. In this paper, we develop a new and bold spectrum-sharing paradigm beyond the state of the art for future wireless networks. We explore network cooperation as a new dimension for spectrum sharing between the primary and secondary users. Such network cooperation can be defined as a set of policies under which different degrees of cooperation are to be achieved. The benefits of this paradigm are numerous, as they allow integrating resources from two networks. There are many possible node-level cooperation policies that one can employ under this paradigm. For the purpose of performance study, we consider a specific policy called *U*nited cooperation of *P*rimary and *S*econdary (UPS) networks. UPS allows a complete cooperation between the primary and secondary networks at the node level to relay each other's traffic. As a case study, we consider a problem with the goal of supporting the rate requirement of the primary network traffic while maximizing the throughput of the secondary sessions. For this problem, we develop an optimization model and formulate a combinatorial optimization problem. We also develop an approximation solution based on a piece-wise linearization technique. Simulation results show that UPS offers significantly better throughput performance than that under the interweave paradigm.

Index Terms—Cognitive radio, node-level cooperation, primary network, secondary network, spectrum sharing

1 Introduction

The last decade has witnessed rapid advance in the research and development of spectrum-sharing technologies. Recent report by the President's Council of Advisors on Science and Technology (PCAST) [16] called for the sharing of 1 GHz of federal government radio spectrum with non-government entities in order to spur economic growth. This report further accelerated the pace of commercialization of innovative spectrum-sharing technologies. As a contributor (J.H. Reed) to the PCAST report, our team began to realize that what was needed was a much more aggressive and broader vision for enhancing spectrum utilization. In [7], Goldsmith et al. outlined three spectrum-sharing paradigms for cognitive radios (CR), namely *underlay*, *overlay*, and *interweave*. These three paradigms were defined from an *information theoretic* perspective, solely based on how much side information (e.g., channel conditions, codebooks) is available to the CRs. In the networking community, these three paradigms have been mapped into specific scenarios of how primary and secondary networks interact with each other for data forwarding. Specifically, the *interweave* paradigm refers to the simple idea that secondary users are allowed to use a spectrum band allocated to the primary users only when the primary users are not using the band [1], [2], [3], [6], [8], [21], [28]. This paradigm is analogous to the classic interference avoidance in medium access, or in CR terminology, dynamic spectrum access (DSA). This is the prevailing scenario on which most of research efforts have been devoted by the CR community in recent years.

The *underlay* paradigm refers to that secondary users' activities or interference on primary users is negligible (or below a given threshold). In contrast to the interweave paradigm, secondary users may be active *concurrently* with the primary users in the same vicinity and in the same frequency. Potential interference from the secondary users may be properly canceled (by the secondary users) via various interference cancelation (IC) techniques so that residual interference are negligible to the primary users [5], [13], [24], [25], [27].

Finally, the *overlay* paradigm requires that the secondary users have the primary users' codebook and messages so that the secondary users can help maintain or improve the communication of the primary users while still achieving some communication on their own. This is accomplished through sophisticated signal processing and coding (e.g., dirty paper coding (DPC) [4], [22] and power allocation [12]). From a networking perspective, the overlay paradigm can be interpreted as having secondary users help forward traffic of the primary users on top of its own communications.

Under the interweave and underlay paradigms, the primary and secondary networks are independent (in terms of

- *X. Yuan, Y. Shi, X. Qin, Y. T. Hou, W. Lou, S. F. Midkiff, and J. H. Reed are with the Virginia Polytechnic Institute and State University, Blacksburg, VA 24061. E-mail: {xuy10, yshi, xiaoqi, thou, wjlou, midkiff, reedjh}@vt.edu.*
- *S. Kompella is with the U.S. Naval Research Laboratory, Washington, DC 20375. E-mail: sastry.kompella@nrl.navy.mil.*

Manuscript received 19 Nov. 2014; revised 16 Dec. 2015; accepted 2 Mar. 2016. Date of publication 7 Mar. 2016; date of current version 1 Dec. 2016.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TMC.2016.2539161

data forwarding in each network). On the other hand, under the overlay paradigm, there is some level of cooperation by the secondary network. Inspired by this primitive cooperation idea in the overlay paradigm, there have been some recent efforts [9], [10], [14], [15], [19], [20], [26] on how to exploit possible cooperation from secondary users for the benefit of data forwarding. We will review these efforts in detail in Section 2. To summarize, the focus of these efforts has been limited to having secondary nodes help relay primary nodes' traffic. This, as we envision in this paper, is only a tip of the iceberg.

In this paper, we develop a paradigm with a much broader vision beyond the state of the art. We explore *network cooperation* as a new dimension for spectrum sharing between primary and secondary nodes. Such network cooperation can be defined as a set of *policies* under which different degrees of cooperation are to be achieved. Corresponding to each cooperation policy, a traffic-forwarding behavior for primary and secondary users can be defined. One such primitive policy, as that in [9], [10], [14], [15], [19], [20], [26], is to have secondary network help relay primary users' traffic. Another policy (United cooperation of Primary and Secondary (UPS) [23]), which we will use as a main policy example in this paper, is to allow complete node-level cooperation between the primary and secondary networks for data forwarding. These two examples are among many possible policies that one can define to achieve network sharing between primary and secondary networks.

To concretize our discussion on policy-based network sharing, we consider the UPS policy in detail, where UPS is the abbreviation of *U*nited cooperation of *P*rimary and *S*econdary networks [23]. UPS represents a policy that allows a complete cooperation between the primary and secondary networks to relay each other's traffic. For performance evaluation, we study a problem with the goal of supporting the rate requirements of the primary sessions while maximizing the throughput of the secondary sessions. A number of technical challenges must be addressed in this problem, including how to provide guaranteed service for the primary traffic while supporting as much the secondary traffic as possible, how to select the optimal relays and routing paths for each source and destination pair, and how to coordinate the transmission and interference relationship between the primary and secondary nodes. For this problem, we develop an optimization model and formulate a combinatorial optimization problem. Although the problem is in the form of mixed-integer nonlinear program (MINLP), we develop an approximation solution based on the piece-wise linearization technique that allows to transform this problem into a mixed-integer linear program (MILP). Through simulation results, we demonstrate that UPS policy offers significantly better throughput performance than that under the existing interweave paradigm.

The remainder of this paper is organized as follows. In Section 2, we review related work on primary and secondary network cooperation. In Section 3, we outline our vision of policy-based network cooperation and use UPS as an example. In Section 4, we use UPS as a case study for performance evaluation. For UPS, we develop an optimization model and formulate an optimization problem. In Section 5, we propose an approximation solution for the UPS throughput optimization problem. Section 6 presents simulation results to demonstrate the benefits and advantages of the UPS policy. Section 7 concludes this paper and points out future research directions.

2 Related Work

Due to space limitation, we will focus our attention on recent research efforts related to primary and secondary network cooperation. We find that all these efforts only considered having the secondary network help relay traffic for the primary network. In [19], Simeone et al. proposed to have the primary network lease its spectrum in the time domain to the secondary network in exchange for having the secondary network relay its data. In [26], Zhang and Zhang formulated this model as a Stackelberg game and a unique Nash Equilibrium point was achieved for maximizing primary and secondary users' utilities in terms of their transmission rates and revenue. In [20], Su et al. proposed to have the primary network lease its spectrum in the frequency domain to the secondary network to relay its data in order to maximize primary users' energy saving and secondary users' data rates. In [10], Jayaweera et al. proposed a new way to encourage primary users to lease their spectrum by having secondary users place bids on the amount of power they are willing to expend for relaying primary users' traffic. In [9], Hua et al. proposed a MIMO-based cooperative CR network where the secondary users utilize MIMO's antenna diversity to help relay primary users' traffic while transmitting their own traffic. In [14], Manna et al. considered the three-node model in [11]. The relay node was assumed to be a secondary node and have MIMO capability. The primary transmitter leases the second time slot to the secondary node (relay node) so that the secondary node can use the time slot to help relay the primary node's traffic while transmitting its own data. In [15], Nadkar et al. considered how to offer incentive (in terms of time and frequency) to a secondary network to help transmit primary user traffic. They studied a cross-layer optimization problem that maximizes transmission opportunities for secondary users while offering a guaranteed throughput to the primary users.

In all these efforts involving node-level cooperation between the primary and secondary networks, the focus has been limited to having secondary nodes help primary nodes in relaying primary users' traffic. As discussed, this is only a tip of the iceberg on network cooperation. In this paper, we envision much broader cooperation between the two networks.

3 Case of Policy-Based Network Cooperation

As discussed in Section 1, the goal of this paper is to outline a broad vision of policy-based network cooperation between the primary and secondary networks as a new dimension in radio spectrum sharing. Here, a policy defines the scope of cooperation at the node-level between the two networks. Such cooperation policies could vary from unilateral cooperation (i.e., only secondary nodes help relay primary user traffic but not vice versa), bilateral cooperation, constrained cooperation, or other customized policy based on particular application needs or requirements.

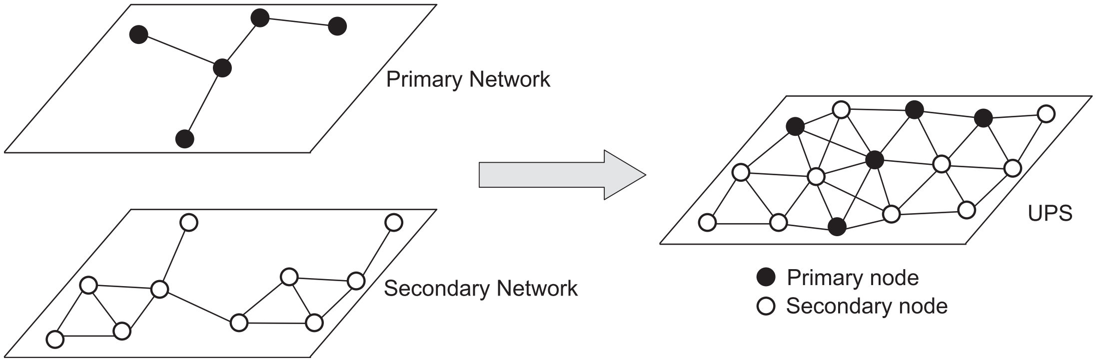


Fig. 1. Network topologies under the interweave and the UPS policy.

As a concrete example, we consider the UPS policy discussed in Section 1, which represents an interesting and extreme scenario where there is complete cooperation between the primary and secondary networks. Fig. 1 illustrates the UPS policy for multi-hop primary and secondary networks. Unlike overlay, which is limited to only allowing secondary nodes help relay primary nodes' traffic, UPS allows primary nodes to help relay secondary nodes' traffic as well. From a network resource perspective, the UPS policy allows the pooling of all the resources from primary and secondary networks together and allows users in each network to access much richer network resources in a combined network. Note that although the two networks are combined into one at the physical level, priority or service guarantee to the primary network traffic can still be enforced by implementing appropriate traffic engineering rules.

It is not hard to see that there are many potential benefits associated with the UPS policy. We briefly describe these benefits as follows:

- *Topology*. Comparing to having primary and secondary nodes being independent for each other, the combined network allows both primary and secondary networks a much improved connectivity with nodes from both networks.
- *Power Control*. As more nodes fall in the maximum transmission range of a primary or secondary node, this node has more flexibility in choosing its next hop node via power control. This flexibility can be exploited for different upper layer performance requirements or objectives.
- *Link Layer*. The improved physical topology allows more opportunities at the link layer for spectrum access. Both the primary and secondary networks can better coordinate with each other in transmission and interference avoidance. Further, the potential issue associated with link failure can now be mitigated effectively.
- *Network Diversity*. The combined network offers more routing opportunities to users in both networks. This directly translates into improved throughput and delay performance for user sessions.
- *Service and Applications*. The UPS architecture (combining both primary and secondary networks) allows to offer much richer services and applications than those services that were studied in [9], [10], [14], [15], [19], [20], [26]. Although the two networks are combined, the services and applications offered to users in each network can still be supported, by implementing certain traffic engineering policies. In other words, the combined network does not mean that service guarantee to the primary network will be lost. On the contrary, by specifying the desired resource management policy appropriately in the combined network, one can easily achieve various service differentiation objectives and application goals, as we shall describe in a case study in the rest of this paper.

The above UPS only represents one policy under the policy-based network cooperation paradigm. There are many other policies that can also be considered, ranging from no cooperation, unilateral cooperation, constrained cooperation, among others. Interweave and UPS can be considered two extreme cases of the policy space for network cooperation. The overlay paradigm that we discussed earlier (i.e., only secondary nodes helping primary traffic but not vice versa) may resemble the unilateral sharing policy, which can be viewed as a policy between the interweave and UPS. The *constrained cooperation* policy allows each network to only engage a subset of its nodes in network cooperation. The motivation of this policy is that certain nodes in either network may be too critical or sensitive (e.g., due to security concerns) in its own network and are thus prohibited from interacting with nodes from the other network. This constrained cooperation may be viewed as a generalization of interweave and UPS. Again, the policies discussed above only represent a few among a lot of possibilities. The definition of a policy is up to the network operators and it determines the scope of cooperation between the two networks.

The policy-based node-level cooperation paradigm may offer many possibilities and potential benefits for both the primary and secondary networks. From a networking perspective, the improved network connectivity, increased flexibility in power control, scheduling and routing all translate into improved forwarding performance for primary and secondary users' traffic. From a spectrum-sharing perspective, the ability to access other network infrastructure helps improve spatial diversity, thus allowing users to tap unused spectrum in the spatial domain. From economic perspective, such shared network infrastructure reduces the cost of infrastructure needed for each individual network (by allowing the tapping of another network's infrastructure resource), thus helping to enable traditionally underserved population and areas to benefit from current and future wireless-enabled goods and services. But from regulatory perspective, the proposed policy-based node-level cooperation paradigm may be ahead of its time. But there is no reason why we should not investigate its capability and recognize its potential from a research perspective. This is the goal of this paper.

4 Case Study: UPS Policy

4.1 Problem Scope

In the rest of this paper, we offer an in-depth study of the UPS policy. Referring to Fig. 1, suppose that there is a set of sessions in the primary network, with each session having a certain rate requirement. In the secondary network, suppose there is also a set of sessions, with each session having an elastic traffic requirement. By "elastic", we mean that each secondary session does not have a stringent rate requirement as the primary session. Instead, each secondary session will be supported on a best-effort basis and will transmit as much as the remaining network resource allows. A plausible goal under the UPS policy could be to have the combined network to support the rate requirements of the primary sessions while maximizing the throughput of the secondary sessions.

For this problem, there are a number of technical challenges that one must address:

- *Guaranteed service for primary traffic*. Since each primary session is assumed to have a hard rate requirement, the combined network should support it at all possibility. This problem alone may not be challenging. What is challenging (and interesting) is that should there are multiple ways to support primary sessions' rate requirements. We should find such a way that the rates for the secondary sessions are maximized in the combined network.
- *Relay selection*. To meet the service requirement (guaranteed service for primary traffic) and to optimize the objective (maximize the rates of secondary sessions), relay node selection along a route (for either a primary or secondary session) is not a trivial problem.
- *Scheduling*. To maximize the rates of the secondary sessions while guaranteeing the rates of the primary sessions, scheduling in each time slot needs to be carefully designed. In particular, in addition to addressing traditional self-interference (half-duplex) and mutual-interference problems, the primary network must be cooperative so as to help the secondary sessions to achieve their optimization objective in the combined network. Such cooperative behavior from the primary network is a key in the UPS policy and has not been explored in prior efforts.

4.2 Mathematical Modeling

In this section, we develop a mathematical model for the UPS policy. Table 1 lists notation in this paper. Denote $\mathcal{N}$ as the combined set of nodes consisting the set of primary nodes $\hat{\mathcal{N}}_{\rm P}$ and the set of secondary nodes $\mathcal{N}_{\rm S}$, i.e., $\mathcal{N} = \hat{\mathcal{N}}_{\rm P} \bigcup \mathcal{N}_{\rm S}$. In the combined network, denote $\mathcal{T}_i$ as the set of nodes (including both primary and secondary nodes) that is located within a nodes i's transmission range, where i can be either a primary or secondary node (i.e., $i \in \mathcal{N}$). Denote $\mathcal{J}_i$ as the set of nodes (including both primary and secondary nodes) that is located within node j's interference range, where j can be either a primary or secondary node. For a primary session $l \in \hat{\mathcal{L}}$, we assume it has a hard requirement on its data rate, which we denote as $\hat{R}(l)$. For a secondary session $m \in \mathcal{L}$, we assume that it does not have a rate requirement. Instead, the data rate $r(m)$ on $m \in \mathcal{L}$ is supported on a best-effort basis and will be an optimization variable in the problem formulation.

TABLE 1
Notation

Primary Network	
$\hat{\mathcal{N}}_{\rm P}$	The set of primary nodes
$\hat{\mathcal{L}}$	The set of primary sessions
$\hat{f}_{ij}(l)$	The flow rate traversing on link (i,j) that is attributed to primary session $l \in \hat{\mathcal{L}}, i, j \in \mathcal{N}$
$\hat{s}(l)$	The source node of primary session $l \in \hat{\mathcal{L}}$
$\hat{d}(l)$	The destination node of primary session $l \in \hat{\mathcal{L}}$
$\hat{R}(l)$	The data rate requirement of primary session $l \in \hat{\mathcal{L}}$
Secondary Network	
$\mathcal{N}_{\rm S}$	The set of secondary nodes
$\mathcal{L}$	The set of secondary sessions
$f_{ij}(m)$	The flow rate traversing on link (i,j) that is attributed to secondary session $m \in \mathcal{L}, i, j \in \mathcal{N}$
$s(m)$	The source node of secondary session $m \in \mathcal{L}$
$d(m)$	The destination node of secondary session $m \in \mathcal{L}$
$r(m)$	The data rate achieved by secondary session $m \in \mathcal{L}$
Combined Network	
$\mathcal{N}$	The set of all nodes in the network, $\mathcal{N} = \hat{\mathcal{N}}_{\rm P} \bigcup \mathcal{N}_{\rm S}$
C_{ij}	The link capacity of link $(i,j), i, j \in \mathcal{N}$
$x_{ij}[t]$	$= 1$ if node i is transmitting data to node j in time slot t, and is 0 otherwise
$\mathcal{T}_i$	The set of nodes that are located within the transmission range of node $i \in \mathcal{N}$
$\mathcal{J}_i$	The set of nodes that are located within the interference range of node $i \in \mathcal{N}$
T	The number of time slots in a frame

Guaranteed service for the primary sessions. For primary sessions, they consider the combined network $\mathcal{N}$ as their communication resources. For flexibility and load balancing, we allow flow splitting in the network. That is, the flow rate of a session may split and merge inside the network in whatever loop-free manner as long as it can help support the given rate requirement $\hat{R}(l)$ of session $l \in \hat{\mathcal{L}}$. Denote $\hat{f}_{ij}(l)$ as the data rate on link (i,j) that is attributed to primary session $l \in \hat{\mathcal{L}}$, where $i \in \mathcal{N}$ and $j \in \mathcal{T}_i$. Denote $\hat{s}(l)$ and $\hat{d}(l)$ as the source and destination nodes of primary session $l \in \hat{\mathcal{L}}$, respectively. We have the following flow balance constraints:

- If node i is the source node of primary session $l \in \hat{\mathcal{L}}$ (i.e., $i = \hat{s}(l)$), then

$$\sum_{j \in \mathcal{T}_i} \hat{f}_{ij}(l) = \hat{R}(l) \qquad (l \in \hat{\mathcal{L}}). \tag{1}$$

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

$$\sum_{j\in\mathcal{T}_i}^{j\neq\hat{s}(l)} \hat{f}_{ij}(l) = \sum_{k\in\mathcal{T}_i}^{k\neq\hat{d}(l)} \hat{f}_{ki}(l) \qquad (l\in\hat{\mathcal{L}}, i\in\hat{\mathcal{N}}_{\mathrm{P}}). \tag{2}$$

- If node i is the destination node of primary session l (i.e., $i=\hat{d}(l)$), then

$$\sum_{k\in\mathcal{T}_i} \hat{f}_{ki}(l) = \hat{R}(l) \qquad (l\in\hat{\mathcal{L}}). \tag{3}$$

It can be easily verified that once (1) and (2) are satisfied, then (3) is also satisfied. As a result, it is sufficient to list only (1) and (2) in the formulation.

Best-effort service for secondary sessions. Under the UPS policy, the primary sessions have priority in access the combined network resources (in the form of guaranteed services). Once the primary sessions are supported, the secondary sessions may use as much as the remaining resources in the combined network. How the primary and secondary sessions interact in the combined network should be part of an optimization problem. Denote $f_{ij}(m)$ as the data rate on link (i,j) that is attributed to secondary session $m\in\mathcal{L}$. Denote $s(m)$ and $d(m)$ as the source and destination nodes of secondary session $m\in\mathcal{L}$, respectively. Similar to that for the primary sessions, we allow flow splitting for the secondary sessions. We have the following flow balance constraints:

- If node i is the source node of secondary session $m\in\mathcal{L}$ (i.e., $i=s(m)$), then we have

$$\sum_{j\in\mathcal{T}_i} f_{ij}(m) = r(m) \qquad (m\in\mathcal{L}). \tag{4}$$

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

$$\sum_{j\in\mathcal{T}_i}^{j\neq s(m)} f_{ij}(m) = \sum_{k\in\mathcal{T}_i}^{k\neq d(m)} f_{ki}(m) \quad (m\in\mathcal{L}, i\in\mathcal{N}_{\mathrm{S}}). \tag{5}$$

- If node i is the destination node of secondary session m (i.e., $i=d(m)$), then

$$\sum_{k\in\mathcal{T}_i} f_{ki}(m) = r(m) \qquad (m\in\mathcal{L}). \tag{6}$$

Again, to avoid redundancy, it is sufficient to list only (4) and (5) in the formulation.

Note that although (4)-(6) are similar to (1)-(3), there is an important difference between them: unlike $\hat{R}(l)$ for primary session $l\in\hat{\mathcal{L}}$, which is a given *constant*, secondary session rate $r(m)$, $m\in\mathcal{L}$, is an optimization *variable*. Therefore, for the primary sessions, we only need to optimize their flow paths, while for the secondary sessions, we need to optimize both their routes and their rates.

Self-interference constraints. We assume scheduling is done in time slot on a frame-by-frame basis, with each frame consisting of T time slots. We use a binary variable $x_{ij}[t], i,j\in\mathcal{N}$ and $1\le t\le T$, to indicate whether node i transmits data to node j. That is,

$$x_{ij}[t] = \begin{cases} 1 & \text{If node } i \text{ transmits data to node } j \\ & \text{in time slot } t; \\ 0 & \text{otherwise,} \end{cases}$$

where $i\in\mathcal{N}, j\in\mathcal{T}_i$, and $1\le t\le T$.

Assuming each primary or secondary session is unicast, a node i only needs to transmit to or receive from one node in a time slot. We have

$$\sum_{j\in\mathcal{T}_i} x_{ij}[t] \le 1 \qquad (i\in\mathcal{N}, 1\le t\le T)\ , \tag{7}$$

$$\sum_{k\in\mathcal{T}_i} x_{ki}[t] \le 1 \qquad (i\in\mathcal{N}, 1\le t\le T)\ . \tag{8}$$

To account for half-duplex at each node i, we have:

$$x_{ij}[t] + x_{ki}[t] \le 1 \qquad (i\in\mathcal{N}, j,k\in\mathcal{T}_i, 1\le t\le T)\ . \tag{9}$$

These three constraints in (7), (8) and (9) can be replaced by the following single constraint,

$$\sum_{j\in\mathcal{T}_i} x_{ij}[t] + \sum_{k\in\mathcal{T}_i} x_{ki}[t] \le 1 \qquad (i\in\mathcal{N}, 1\le t\le T). \tag{10}$$

To see this, note that in (10), if node i is receiving data from some node in $\mathcal{T}_i$ in time slot t, we must have $\sum_{j\in\mathcal{T}_i} x_{ij}[t]=0$, i.e., node i cannot transmit in the same time slot. This is exactly the half-duplex constraint. In this case, (10) also becomes (8). On the other hand, if node i is transmitting to some node in $\mathcal{T}_i$ in time slot t, then $\sum_{k\in\mathcal{T}_i} x_{ki}[t]=0$, i.e., node i cannot receive in the same time slot. Again, this is the half-duplex constraint. In this case, (10) becomes (7).

Mutual interference constraints. To model mutual interference constraints, we assume that for any primary or secondary node $j\in\mathcal{N}$ that is receiving data in time slot t, it shall not be interfered by another (unintended) transmitting node $p\in\mathcal{I}_j$ in the same time slot. We have the following mutual interference constraint:

$$x_{ij}[t] + x_{pk}[t] \le 1\ , \tag{11}$$

where $i\in\mathcal{T}_j, p\in\mathcal{J}_j, k\in\mathcal{T}_p, j\in\mathcal{N}, j\neq k$, and $1\le t\le T$.

Following the same token in (10), the three constraints in (7), (8) and (11) can be replaced by the following single and equivalent constraint,

$$\sum_{i\in\mathcal{T}_j} x_{ij}[t] + \sum_{k\in\mathcal{T}_p} x_{pk}[t] \le 1\ , \tag{12}$$

where $p\in\mathcal{J}_j, j\in\mathcal{N}, j\neq k$, and $1\le t\le T$.

Link rate constraints. For each link (i,j), denote the link capacity as C_{ij}, e.g., $C_{ij}=B\log_2(1+\frac{Q_i d_{ij}^{-\alpha}\lambda}{N_0})$, where B is bandwidth, Q_i is the power spectral density from transmit node i, d_{ij} is the distance between node i and j, α is the path loss index, λ is the antenna related constant, and N_0 is the ambient Gaussian power spectral density. Since the aggregate flow rate from the primary and secondary sessions on each link (i,j) cannot exceed the average link rate (over T

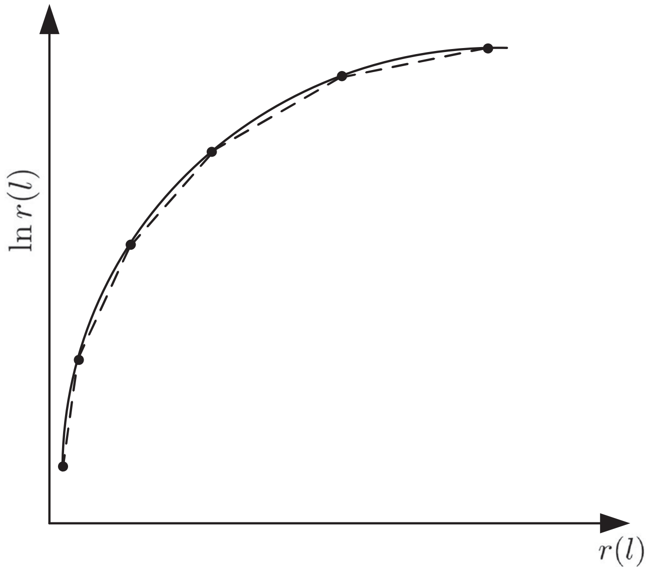


Fig. 2. Piece-wise approximation with line segments.

time slots), we have

$$\sum_{l\in\hat{\mathcal{L}}}^{j\neq\hat{s}(l),i\neq\hat{d}(l)} \hat{f}_{ij}(l) + \sum_{m\in\mathcal{L}}^{j\neq s(m),i\neq d(m)} f_{ij}(m) \leq \frac{1}{T}\sum_{t=1}^{T} C_{ij}\cdot x_{ij}[t]. \quad (13)$$

4.3 Problem Formulation

In the combined network, our goal is to offer guaranteed support for the primary sessions (each with a given rate requirement) while maximizing the throughput for the secondary sessions, whose traffic is assumed to be elastic. For maximizing secondary network throughput, different objective functions can be explored to satisfy network requirement. In [23], we considered a simple case with linear objective function (i.e., maximizing the minimum throughput). In this paper, we will consider a nonlinear objective function. We use a utility function $\ln r(m)$ for $m \in \mathcal{L}$ as our objective. Such utility function is widely used in the literature [17]. We have the following problem formulation:

OPT
max $\sum_{m\in\mathcal{L}} \ln r(m)$
s.t. Guaranteed service for primary sessions: (1), (2);
Best effort service for secondary sessions: (4), (5);
Self interference constraints: (10);
Mutual interference constraints: (12);
Link capacity constraints: (13).

In this formulation, $\hat{R}(l)$ and C_{ij} are constants, $x_{ij}[t]$ are binary variables, $\hat{f}_{ij}(l)$, $f_{ij}(m)$ and $r(m)$ are continuous variables. Due to nonlinear terms $\ln r(m)$ in the objective function and binary variables $x_{ij}[t]$, the optimization problem is a *mixed-integer nonlinear programming*, which is NP-hard in general. In the next section, we develop an approximation algorithm to solve this problem.

5 An Approximate Solution

5.1 Overview

In this section, we develop an approximate solution to OPT with guaranteed performance. For the nonlinear *log* term in the objective function of OPT, one could relax the nonlinear function with a series of linear functions. The issue here is how to achieve such linearization with performance guarantee. This is the focus of our proposed solution.

For a target performance gap ϵ between the optimal objective (unknown) and the approximate objective (that we aim to develop), we will develop an algorithm to determine a set of piece-wise linear segments that approximate the log function (See Fig. 2). The essence of our linear approximation is to find just the right number of linear and unequal-length segments to approximate the log function. The idea is that, for a given performance gap ϵ, we can calculate the maximum linear approximation error, say η, that is allowed in the linearization. Then, we can develop an algorithm (Section 5.2) to find the slopes and starting points for the set of linear segments. Subsequently, the nonlinear log terms in OPT can be replaced by a set of linear constraints and we have a new linearized optimization problem, which we denote as OPT-L. Although OPT-L is in the form of *mixed-integer linear programming*, the integer variables are all binary. We find that commercial software (such as CPLEX) can solve such binary MILP efficiently.

5.2 Linearization

Our goal of linear approximation of $\ln r(m)$ is to replace $\ln r(m)$ with the minimum number of linear segments while ensuring that the difference between any point on $\ln r(m)$ and its corresponding linear segment is no more than η. Denote K_m as the minimum number of line segments such that each segment meets the error requirement (i.e., η). Denote $r_L(m)$ and $r_U(m)$ as the lower and upper bounds for $r(m)$, respectively. For $r_L(m)$, we can set it to an arbitrarily small positive value. For $r_U(m)$, we can set it to $\max_{i,j\in\mathcal{N}} C_{ij}$, the maximum capacity among all links. Denote $r_0(m), r_1(m), \ldots, r_{K_m}(m)$ as values on the X-axis for the end points of these K_m segments, with $r_0(m) = r_L(m)$ and $r_{K_m}(m) = r_U(m)$.

The minimum number of line segments K_m can be found with the following iterative process. We start from $r_0(m)$ to calculate the slope of the first segment, which must ensure that this segment satisfies the error bound η. After finding this slope, we can find the right-side end point of the first segment. From this point, we repeat the same process for the second segment and so forth, until the last segment exceeds $r_U(m)$.

Specifically, denote slope of the kth linear segment as $q_k(m)$, i.e.,

$$q_k(m) = \frac{\ln r_k(m) - \ln r_{k-1}(m)}{r_k(m) - r_{k-1}(m)}. \quad (14)$$

Denote $y_k(r(m))$ as the kth linear segment that approximates $\ln r(m)$. Then we have:

$$y_k(r(m)) = q_k(m)\cdot[r(m) - r_{k-1}(m)] + \ln r_{k-1}(m), \quad (15)$$

for $r_{k-1}(m) \leq r(m) \leq r_k(m)$.

Referring to Fig. 3, for any point $r(m)$ within $r_{k-1}(m) \leq r(m) \leq r_k(m)$, it is easy to see that the point on the tangential line (in parallel to the linear segment approximation) that intersects the $\log$ curve has the maximum approximation error η. Denote the X-coordinate of this point

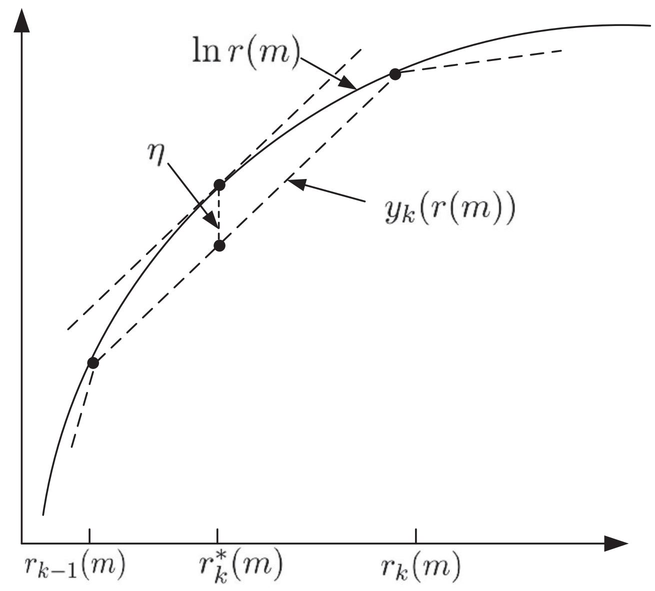


Fig. 3. An illustration of the maximum approximation error for piece-wise line segment.

as $r_k^*(m)$, we have $\eta = \ln r_k^*(m) - y_k(r_k^*(m))$. Since the slope of tangential line (achieving η) for $\ln r(m)$ is $\frac{1}{r(m)}$, then $q_k(m) = \frac{1}{r_k^*(m)}$, or $r_k^*(m) = \frac{1}{q_k(m)}$, where $q_k(m)$ is the slope of the linear segment $y_k(r(m))$. Then, we have

$$\begin{aligned}\eta &= \ln r_k^*(m) - y_k(r_k^*(m))\\ &= \ln r_k^*(m) - \Big[q_k(m)\cdot\Big(r_k^*(m) - r_{k-1}(m)\Big) + \ln r_{k-1}(m)\Big]\\ &= \ln\frac{1}{q_k(m)} - q_k(m)\cdot\left(\frac{1}{q_k(m)} - r_{k-1}(m)\right) - \ln r_{k-1}(m)\\ &= -\ln q_k(m) - 1 + q_k(m)\cdot r_{k-1}(m) - \ln r_{k-1}(m).\end{aligned}$$

Therefore, we have the following equation:

$$-\ln q_k(m) + q_k(m)\cdot r_{k-1}(m) - [\ln r_{k-1}(m) + \eta + 1] = 0. \tag{16}$$

For a give error bound η, the values of $r_1(m),\ldots,r_{K_m}(m)$ and slopes $q_1(m), q_2(m),\ldots,q_{K_m}(m)$ can be found iteratively through the following algorithm:

Algorithm 1. (Piece-wise linearization)
Initialization: $k := 0$ and $r_k(m) := r_L(m)$.
While ($r_k(m) < r_U(m)$) {
$k := k+1$.
Find slope $q_k(m)$ by solving the equation (16).
With $q_k(m)$, compute $r_k(m)$ via (14). }
$K_m := k, r_{K_m}(m) := r_U(m)$.
Recalculate $q_{K_m}(m)$ with (14).

The values of $q_k(m)$ in (16) and $r_k(m)$ in (14) can be solved by numerical methods such as bisection method or Newton's method [18].

Lemma 1. *The maximum approximation error within each linear segment as defined by Algorithm 1 is no more than* η.

Proof. The proof is based on the above construction. We omit its discussion here to conserve space. □

5.3 Approximation Gap

By using the piece-wise linearization algorithm (Algorithm 1), we can approximate the log term $\ln r(m)$ with a series of linear segments, each with an approximation error no more than η. For $r(m)$, denote $y(m)$ as the concatenation of the piece-wise linear segments constructed by Algorithm 1. Then the objective function $\max\sum_{m\in\mathcal{L}}\ln r(m)$ in OPT is replaced by the following linear objective and a set of linear constraints (representing the convex hull below the linear segments):

$$\max \quad \sum_{m\in\mathcal{L}} y(m) \tag{17}$$

$$\begin{aligned}\text{s.t.}\quad & y(m) \le q_k(m)\cdot(r(m) - r_{k-1}(m)) + \ln r_{k-1}(m)\\ & (k = 1,2,\ldots,K_m, m\in\mathcal{L}).\end{aligned} \tag{18}$$

The original OPT can be re-formulated into a new optimization problem, which we denote as OPT-L.

$$\begin{aligned}&\text{OPT-L}\\ &\quad\max\quad \textstyle\sum_{m\in\mathcal{L}} y(m)\\ &\quad\text{s.t.}\quad \text{Constraints (1), (2), (4), (5), (10), (12), (13), (18),}\\ &\qquad\qquad r_L(m) \le r(m) \le r_U(m), (m\in\mathcal{L})\\ &\qquad\qquad x_{ij}[t]\in\{0,1\}, f_{ij}(m)\ge 0, \hat{f}_{ij}(l)\ge 0.\\ &\qquad\qquad (i\in\mathcal{N}, j\in\mathcal{T}_i, m\in\mathcal{L}, l\in\hat{\mathcal{L}}, 1\le t\le T),\end{aligned}$$

where $x_{ij}[t]$ are binary variables, $\hat{f}_{ij}(l), f_{ij}(m), r(m)$ and $y(m)$ are continuous variables, and $q_k(m), r_{k-1}(m)$ and $\hat{R}(l)$ are constants. OPT-L is in the form of *mixed integer linear programming*. Since all integers are binary, the MILP problem tends to be solved efficiently by a commercial solver (CPLEX). Our simulation results in Section 6 confirm that this is indeed the case.

We now quantify the gap between the optimal objective values of OPT-L and OPT.

Lemma 2. *The gap between the optimal objective values of OPT and OPT-L,* ϵ, *is upper bounded by* $|\mathcal{L}|\cdot\eta$.

Proof. Suppose an optimal solution of OPT is $\varphi^*_{\text{OPT}} = [x^*_{ij}[t], r^*(m), f^*_{ij}(m), \hat{f}^*_{ij}(l)]$, with the objective value being $Y^*_{\text{OPT}} = \sum_{m\in\mathcal{L}}\ln r^*(m)$. We can construct a feasible solution to OPT-L, denoted as $\varphi_{\text{OPT-L}}$, based on φ^*_{OPT} as follows: $\varphi_{\text{OPT-L}} = [x_{ij}[t], r(m), f_{ij}(m), \hat{f}_{ij}(l), y(m)]$, where $x_{ij}[t] = x^*_{ij}[t], r(m) = r^*(m), f_{ij}(m) = f^*_{ij}(m)$ and $\hat{f}_{ij}(l) = \hat{f}^*_{ij}(l)$. Then, $\varphi_{\text{OPT-L}}$ satisfy constraints (1), (2), (4), (5), (10), (12), (13) in OPT-L. $y(m)$ can be calculated by solving OPT-L with the variables being set to those values in $\varphi_{\text{OPT-L}}$. Suppose that $r^*(m)$ falls in the interval $[r_{k-1}(m), r_k(m)]$. Then the objective function $\sum_{m\in\mathcal{L}} y(m)$ is maximized only when $y(m) = y_k(r^*(m)) = q_k(m)\cdot(r^*(m) - r_{k-1}(m)) + \ln r_{k-1}(m)$ in (18). Denote this objective value in OPT-L as $Y_{\text{OPT-L}}$. Then,

$$\begin{aligned}Y^*_{\text{OPT}} - Y_{\text{OPT-L}} &= \sum_{m\in\mathcal{L}}\ln r^*(m) - \sum_{m\in\mathcal{L}} y(m)\\ &= \sum_{m\in\mathcal{L}}\ln r^*(m) - \sum_{m\in\mathcal{L}} y_k(r^*(m))\\ &= \sum_{m\in\mathcal{L}}\big[\ln r^*(m) - y_k(r^*(m))\big]\\ &\le |\mathcal{L}|\cdot\eta,\end{aligned}$$

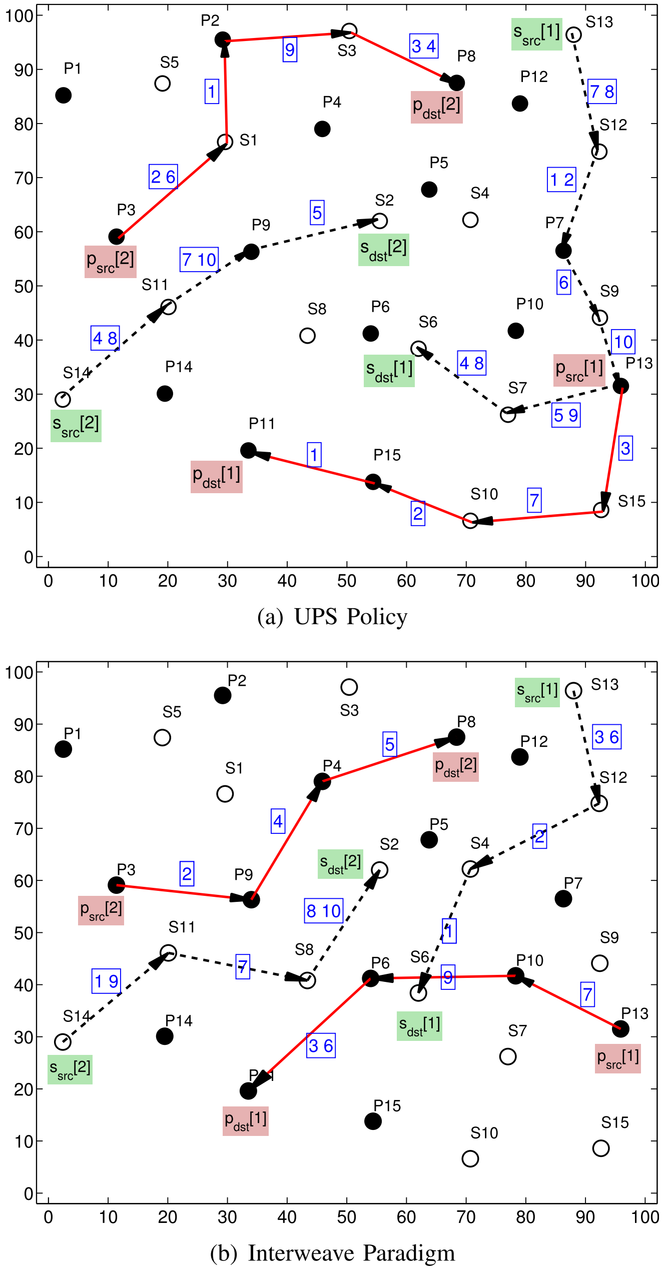


Fig. 4. Region 1 example that showing the flow routing topologies and scheduling for the primary and secondary sessions, where the solid line segments are for the primary sessions while the dashed line segments are for the secondary sessions.

where last inequality holds by Lemma 1. We let $\epsilon = |\mathcal{L}| \cdot \eta$.

Now denote $\varphi^*_{\text{OPT-L}}$ as an optimal solution for OPT-L, with the objective value of $Y^*_{\text{OPT-L}}$. Since $Y_{\text{OPT-L}}$ is merely the objective value of a feasible solution, we have $Y^*_{\text{OPT-L}} \geq Y_{\text{OPT-L}}$. Then $Y^*_{\text{OPT}} - Y^*_{\text{OPT-L}} \leq Y^*_{\text{OPT}} - Y_{\text{OPT-L}} \leq \epsilon$. This completes the proof. □

Our complete solution for solving OPT can be summarized as follows: For any a given performance gap ϵ, we can compute linear approximation error $\eta = \frac{\epsilon}{|\mathcal{L}|}$. Based on the approximation error η, we perform piece-wise linear approximation through Algorithm 1. Then we reformulate OPT to OPT-L, and solve it by CPLEX.

TABLE 2
Location of Primary and Secondary Nodes for the 30-Node Network

Primary Node	Location	Secondary Node	Location
P_1	(2.5, 85.2)	S_1	(29.6, 76.6)
P_2	(29.2, 95.5)	S_2	(55.5, 62)
P_3	(11.4, 59.1)	S_3	(50.4, 97.1)
P_4	(45.9, 79)	S_4	(70.7, 62.2)
P_5	(63.8, 67.8)	S_5	(19.1, 87.4)
P_6	(54, 41.2)	S_6	(62, 38.4)
P_7	(86.3, 56.5)	S_7	(77, 26.2)
P_8	(68.4, 87.5)	S_8	(43.4, 40.8)
P_9	(34, 56.3)	S_9	(92.4, 44.1)
P_{10}	(78.3, 41.7)	S_{10}	(70.7, 6.6))
P_{11}	(33.5, 19.6)	S_{11}	(20.1, 46.1)
P_{12}	(79, 83.7)	S_{12}	(92.3, 74.8)
P_{13}	(95.9, 31.5)	S_{13}	(88, 96.4)
P_{14}	(19.5, 30.1)	S_{14}	(2.4, 29)
P_{15}	(54.4, 13.8)	S_{15}	(92.6, 8.6)

6 Simulation Results

In this section, we present numerical results to demonstrate the capabilities and advantages of the UPS policy. The goal of this section is twofold. First, we show that the UPS policy offers much better performance for both the primary and secondary networks than that under the interweave paradigm. Second, we shall have a close look at how the primary and secondary nodes help each other in the UPS policy.

6.1 Simulation Setting

We consider a UPS network where both the primary and the secondary nodes are randomly deployed in a 100×100 area. For generality, we normalize the units for distance, bandwidth, power and data rate with appropriate dimensions. We assume the bandwidth of the channel allocated to the primary network is $B = 10$. The number of time slots in a frame is $T = 10$. The transmission power spectral density Q_i for each node $i \in \mathcal{N}$ is 1, the path loss index is 4, the antenna related constant λ is 1, and the ambient Gaussian power spectral density $N_0 = 10^{-6}$. We assume the transmission range and interference range at all nodes are 30 and 50, respectively.

We set the maximum acceptable performance gap between the objective of OPT and its linear approximation OPT-L as $\epsilon = 0.02$.

6.2 An Example

We consider a 30-node network, with 15 primary nodes and 15 secondary nodes randomly deployed in a 100×100 area (see Fig. 4). The location of each node is given in Table 2. In this example, we assume that there are two primary sessions in the primary network and two secondary sessions in the secondary network. The source and destination nodes for each session are randomly chosen in each network and are shown in Table 3. Denote the rate requirements of the two primary sessions as $\hat{R}(1)$ and $\hat{R}(2)$, respectively. We gradually increase the rate requirements of $\hat{R}(1)$ and $\hat{R}(2)$ and examine (i) whether such rates can be supported under the UPS policy and the interweave paradigm, respectively, and (ii) the objective values of secondary session utilities under both the UPS policy and the interweave paradigm.

TABLE 3
Source and Destination Nodes for Each Session in the 30-Node Network

Session	Source	Destination
Primary session 1	P_{13}	P_{11}
Primary session 2	P_3	P_8
Secondary session 1	S_{13}	S_6
Secondary session 2	S_{14}	S_2

The utility maximization problem for the secondary sessions under the interweave paradigm can be formulated following a similar token to OPT.

Table 4 shows the approximation gap between the utility objective of the linearized problem and the utility objective of the original problem under different $\hat{R}(1)$ and $\hat{R}(2)$. The first column represents increasing rate requirements for the primary sessions. The second column shows the utility objectives of the two secondary sessions (abbreviated as "SS" in the table) from the linearized problem OPT-L, while the third column shows the utility objectives of the two secondary sessions from the original problem. The fourth column shows the gap between the utility objectives from the linearized problem and the original problem. Given the target approximation error $\varepsilon = 0.02$, all actual approximation errors fall below this target.

Table 5 summarizes the results of this study. The second column represents increasing rate requirements for the primary sessions (i.e., $\hat{R}(1) = \hat{R}(2)$). For ease of explanation, we break this table into five regions, with each region representing a specific behavior for comparison between the UPS policy and interweave paradigm. The third and fourth columns show the performance under the UPS policy. Specifically, the third column shows whether the rate requirements of the two primary sessions can be supported ("feasible") in the primary network (abbreviated as "PN" in the table); the fourth column shows the rate utility objective of the two secondary sessions (abbreviated as "SS" in the table) with $-\infty$ indicating zero rates for the secondary sessions (due to the log function) and "N/A" indicating not applicable as the corresponding network cannot even support the rate requirements of the primary sessions. The fifth and sixth columns show the performance under the interweave paradigm, which are to be compared to the third and fourth columns under the UPS policy, respectively.

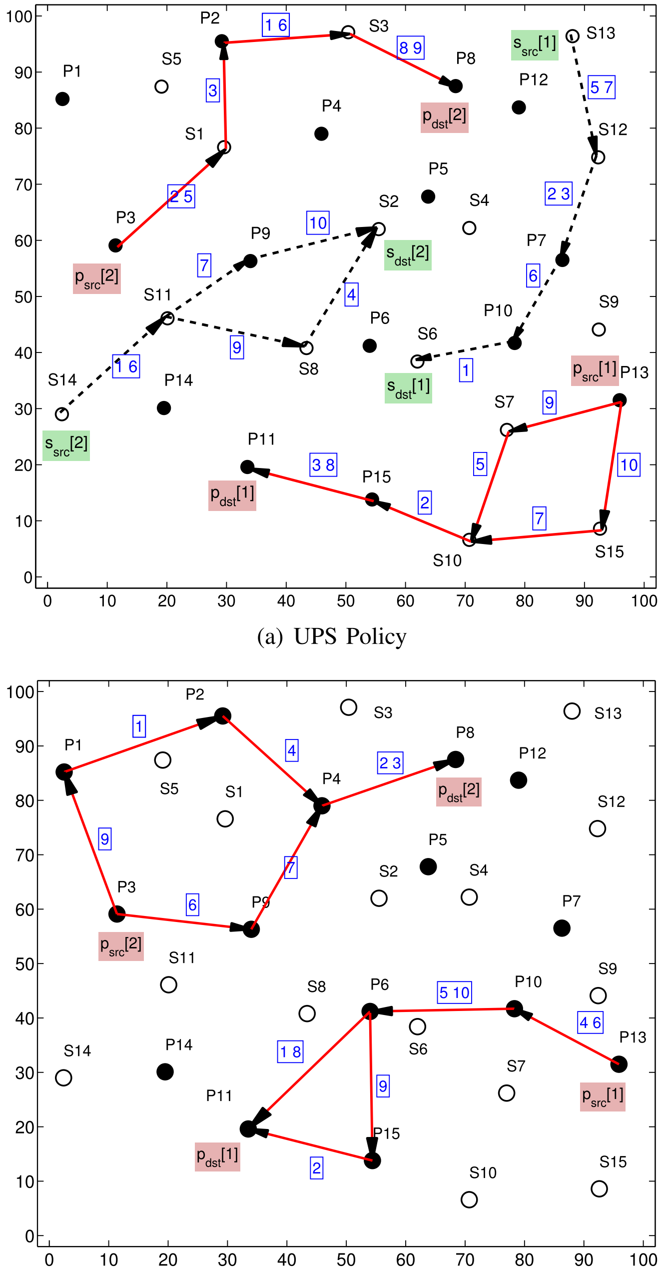
TABLE 4
Approximation Gap Between the SS Utility Objectives of Linearized Problem and Original Problem

Rate Requirement $\hat{R}(1), \hat{R}(2)$	SS Utility of Linearized Problem	SS Utility of Original Problem	Gap
0	3.7012	3.7128	0.0016
0.2	3.288	3.3046	0.0016
0.4	3.288	3.3046	0.0016
0.6	3.288	3.3046	0.0016
0.8	3.288	3.3046	0.0016
1.0	3.288	3.3046	0.0016
1.2	3.288	3.3046	0.0016
1.4	3.288	3.3046	0.0016
1.6	3.288	3.3046	0.0016
1.8	3.158	3.167	0.0009
2.0	3.158	3.167	0.0009
2.2	3.158	3.167	0.0009
2.4	3.158	3.167	0.0009
2.6	2.892	2.899	0.007
2.8	2.653	2.656	0.003
3.0	2.653	2.656	0.003
3.2	2.653	2.656	0.003
3.4	2.653	2.656	0.003
3.6	2.653	2.656	0.003
3.8	2.653	2.656	0.003
4.0	2.288	2.305	0.017
4.2	2.288	2.305	0.017
4.4	2.183	2.191	0.008
4.6	1.969	1.981	0.012
4.8	1.969	1.981	0.012

TABLE 5
Performance Comparison Between the UPS Policy and the Interweave Paradigms for Different Primary Session Rate Requirements

	Rate Requirements	UPS		Interweave Paradigm	
	$\hat{R}(1), \hat{R}(2)$	Feasible in PN	SS Utility	Feasible in PN	SS Utility
Region 1	0	Yes	3.7012	Yes	3.0402
	0.2	Yes	3.288	Yes	1.899
	0.4	Yes	3.288	Yes	1.899
	0.6	Yes	3.288	Yes	1.899
	0.8	Yes	3.288	Yes	1.899
	1.0	Yes	3.288	Yes	1.899
	1.2	Yes	3.288	Yes	1.263
	1.4	Yes	3.288	Yes	1.263
	1.6	Yes	3.288	Yes	1.263
	1.8	Yes	3.158	Yes	1.425
Region 2	2.0	Yes	3.158	Yes	$-\infty$
	2.2	Yes	3.158	Yes	$-\infty$
	2.4	Yes	3.158	Yes	$-\infty$
	2.6	Yes	2.892	Yes	$-\infty$
	2.8	Yes	2.653	Yes	$-\infty$
	3.0	Yes	2.653	Yes	$-\infty$
	3.2	Yes	2.653	Yes	$-\infty$
	3.4	Yes	2.653	Yes	$-\infty$
	3.6	Yes	2.653	Yes	$-\infty$
	3.8	Yes	2.653	Yes	$-\infty$
Region 3	4.0	Yes	2.288	No	N/A
	4.2	Yes	2.288	No	N/A
	4.4	Yes	2.183	No	N/A
	4.6	Yes	1.969	No	N/A
	4.8	Yes	1.969	No	N/A
Region 4	5.0	Yes	$-\infty$	No	N/A
	5.2	Yes	$-\infty$	No	N/A
	5.4	Yes	$-\infty$	No	N/A
	5.6	Yes	$-\infty$	No	N/A
	5.8	Yes	$-\infty$	No	N/A
	6.0	Yes	$-\infty$	No	N/A
	6.2	Yes	$-\infty$	No	N/A
	6.4	Yes	$-\infty$	No	N/A
	6.6	Yes	$-\infty$	No	N/A
	6.8	Yes	$-\infty$	No	N/A
Region 5	7.0	No	N/A	No	N/A

Region 1. This region represents the scenario where the rate requirements of the primary sessions can be supported under both the UPS policy and the interweave paradigm, and the rates of the secondary sessions are positive.



(a) UPS Policy

(b) Interweave Paradigm

Fig. 5. Region 2 example that showing the flow routing topologies and scheduling for the primary and secondary sessions.

Comparing columns four and six, we can find that the secondary sessions always achieve higher utility objectives under the UPS policy than that under the interweave paradigm. This confirms our expectation that the UPS policy can offer higher throughput for the secondary sessions.

As an example, consider the case when both the two primary sessions have rate requirements 1.6. The utility objectives achieved for the secondary sessions under the UPS policy and the interweave paradigms are 3.288 and 1.263, respectively. Specifically, the rates for the two secondary sessions are 4.784 and 5.692 under the UPS policy while the rates for the same two secondary sessions are 1.776 and 2.024 under the interweave paradigm. Under the UPS policy, the flow routing and scheduling for the primary and secondary sessions are shown in Fig. 4a. The number in the box on each link represents the active time slots for this link. Note that primary nodes P_7, P_9 and P_{13} are helping relay secondary sessions' data while secondary nodes S_1, S_3, S_{10} and S_{15} are helping relay the primary sessions' data. In comparison, under the interweave paradigm, the flow routing and scheduling for the primary network is shown in Fig. 4b. According to the time slots used by the primary network, the secondary network calculates the remaining time slots at each node and uses them to maximize their rate utilities. The flow routing and scheduling for the secondary sessions under the interweave paradigm are also shown in Fig. 4b. As expected, there is no cooperation at the node level between the two networks in terms of relaying each other's data.

Region 2. This region represents the scenario where the rate requirements of the primary sessions can be supported under both the UPS policy and the interweave paradigm, while the secondary sessions can only be supported under the UPS policy but not under the interweave paradigm (with zero rate for some sessions and thus $-\infty$ rate utility). This region contains that the combined network can offer more to the secondary sessions than the isolated networks under the interweave paradigm.

As an example, consider the case when the two primary sessions have rate requirements 3.0. The utility achieved for the secondary sessions under the UPS policy is 2.653. Specifically, the rates for the two secondary sessions are 3.753 and 2.793, respectively. Under the UPS policy, the flow routing and scheduling for the primary and secondary sessions are shown in Fig. 5a. Note that primary nodes P_7, P_9, and P_{10} are helping relay secondary sessions' data while secondary nodes S_1, S_3, S_7, S_{10} and S_{15} are helping relay the primary sessions' data. Under the interweave paradigm, the flow routing and scheduling for primary network are shown in Fig. 5b. Based on the time slots used by the primary network, the remaining time slots are not enough to support the secondary sessions, resulting in at least one of the secondary sessions with zero rate. Therefore, the rate utility for the secondary sessions is $-\infty$ under the interweave paradigm.

Region 3. This region represents the scenario where the rate requirements of the primary sessions can be supported under the UPS policy but not so under the interweave paradigm. For the secondary sessions, there is still remaining resource to support them under the UPS policy. For fairness in comparison, we do not consider the rate utilities of the secondary sessions under the interweave paradigm (marked as "N/A"). Region 3 shows the definitive advantage of using a combined network from the primary sessions' perspective over the interweave paradigm.

As an example, we consider the case when the two primary sessions have rate requirements 4.2. The utility objectives achieved by secondary sessions are 2.288 under the UPS policy. Specially, the rates for the two secondary sessions are 3.047 and 3.289, respectively. Under the UPS policy, the flow routing and scheduling for primary and secondary sessions are shown in Fig. 6. Note that primary nodes P_7 and P_{14} are helping relay secondary sessions' data while secondary nodes S_3, S_5, S_{10} and S_{15} are helping relay the primary session' data.

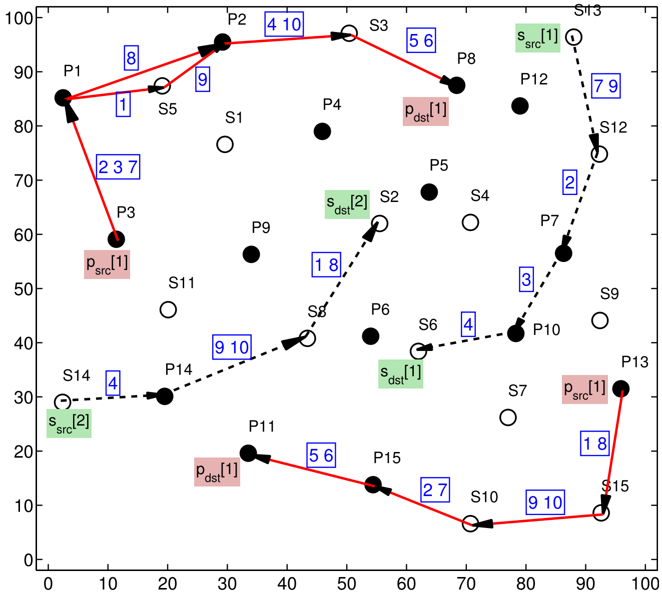


Fig. 6. Region 3 example that showing the flow routing topologies and scheduling for the primary and secondary sessions in the UPS policy.

Region 4. This region represents the scenario where the rate requirement of the primary sessions can be satisfied under the UPS policy but not so under the interweave paradigm. The secondary sessions can no longer be supported under the UPS policy (with zero rate for at least one session and thus $-\infty$ rate utility). For fairness in comparison, we do not consider the rate utilities of the secondary sessions under the interweave paradigm (marked as "N/A") as even the rate requirements for the primary sessions cannot be supported. Similar to Region 3, this region shows the advantage of using a combined network to support the primary sessions over the interweave paradigm

Region 5. As the rate requirements of the primary sessions continue to increase, even the UPS policy will no longer be able to support them after certain point. This is shown in Region 5.

TABLE 6
Average SS Utility Objectives for Different K Users

User Number K	SS utility objectives
5	$-\infty$
10	$-\infty$
15	1.9293
20	2.6653
25	3.2231
30	3.4772

6.3 Varying the Number of Nodes

In this section, we assume there are two primary sessions in the primary network and two secondary sessions in the secondary network. We fix the locations of source and destination nodes of the primary and secondary sessions as shown in Fig. 7. Then, we increase the number of primary and secondary nodes (K) in the network, and these nodes are uniformly deployed in the 100×100 area. Since these additional primary and secondary nodes only serve as relay nodes under UPS, there are no distinction between the two types of nodes.

Table 6 shows the average SS utility objective (over 100 network instances) under different number of nodes (K) in the primary and secondary networks for the case when $\hat{R}(1) = 1.0$ and $\hat{R}(2) = 1.0$. When $K = 5$ and $K = 10$, the network is not dense enough and is not entirely connected. Therefore, the SS utility objectives are both $-\infty$ (i.e., the achievable secondary sessions rate is 0 in both cases). When $K = 15, 20, 25, 30$, the average SS utility objectives increase with the number of users K.

Then we vary $\hat{R}(1)$ and $\hat{R}(2)$ under different network size K. Fig. 8 shows the SS utility objectives under different network size K when $\hat{R}(1)$ and $\hat{R}(2)$ vary. Again, for a given rate for $\hat{R}(1)$ and $\hat{R}(2)$, we have higher SS utility objectives under larger values of K.

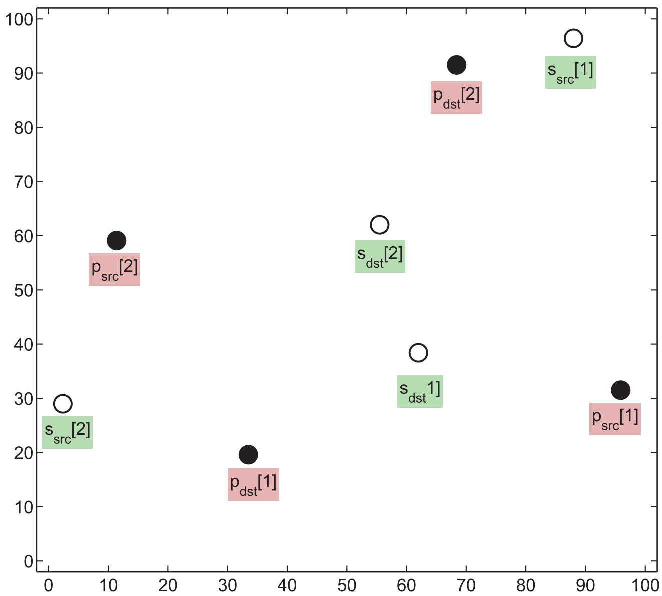


Fig. 7. Locations of the source and destination nodes of the primary and secondary sessions.

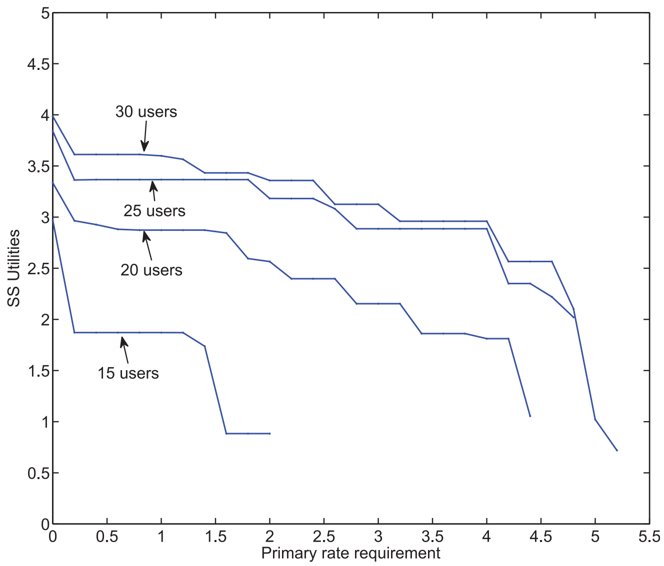


Fig. 8. Comparison of the SS utility objectives for different number of nodes ($K = 10, 15, 20, 25,$ and 30) with the increasing rate requirements for the primary sessions.

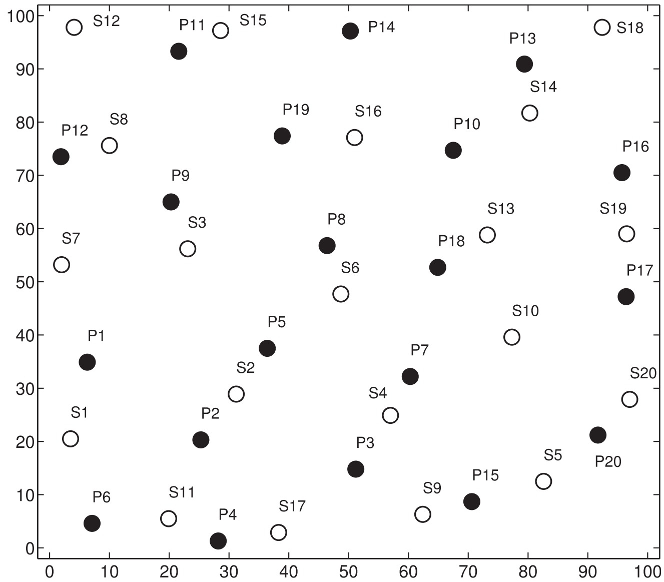


Fig. 9. A 20-node primary network and a 20-node secondary network.

6.4 Varying Session Numbers

In this section, we vary the primary and secondary session numbers. We randomly generate a 20-node primary network and a 20-node secondary network as shown in Fig. 9. In the first part, we will keep the number of secondary sessions fixed and vary the number of primary sessions. In the second part, we will do the converse, i.e., keep the number of primary sessions fixed and varying the number of secondary sessions. In both parts, we will compare the performance under the UPS policy and the interweave paradigm.

Varying the number of primary sessions. Suppose that there are two secondary sessions, with each session's source and destination nodes being (S_{11}, S_7) and (S_4, S_1), respectively. By keeping these secondary sessions fixed, we increase the number of primary sessions. The source and destination nodes of each additional primary session is randomly chosen from the remaining primary nodes. Once chosen, we assume it has a data rate requirement of 1.8 and is added on top of the existing primary sessions. Table 7 shows our results. The first column in the table shows the increasing number of the primary sessions. The second and fourth columns show whether the additional new primary session can be accommodated (feasible) under UPS and interweave, respectively. Comparing these two columns, we can find that the maximum number of the primary sessions under UPS (7) is larger than that under interweave (5). The third and fifth columns show the utility function of the secondary sessions under UPS and interweave. Comparing these two columns, we can see that UPS achieves higher utility objectives than interweave. In summary, both primary and secondary sessions benefit more from UPS than interweave.

TABLE 7
Feasibility Performance of the Primary Sessions and Utilities of the Secondary Sessions Under Increasing Number of the Primary Sessions

Number Of Primary Session	UPS		Interweave Paradigm	
	Feasible in PN	Secondary Utility	Feasible in PN	Secondary Utility
0	Yes	3.69	Yes	2.219
1	Yes	3.446	Yes	1.693
2	Yes	3.058	Yes	0.931
3	Yes	2.661	Yes	0.805
4	Yes	2.118	Yes	$-\infty$
5	Yes	0.83	Yes	$-\infty$
6	Yes	$-\infty$	No	N/A
7	Yes	$-\infty$	No	N/A
8	No	N/A	No	N/A

TABLE 8
Secondary Sessions' Utility Values Under Increasing Number of the Secondary Sessions

Number of Secondary Session	UPS	Interweave
1	2.228	1.355
2	3.21	2.852
3	4.594	2.652
4	4.738	0.943
5	4.253	$-\infty$
6	2.418	$-\infty$
7	2.307	$-\infty$
8	1.134	$-\infty$
9	$-\infty$	$-\infty$

Varying the number of secondary sessions. Now we do the converse. Suppose there are two primary sessions, with each session's source and destination nodes being (P_9, P_{17}) and (P_1, P_{15}), respectively. The data rate requirement for each primary session is 1.8. By keeping these primary sessions fixed, we increase the number of secondary sessions. The source and destination nodes of each additional secondary session is randomly chosen from the remaining secondary nodes. Once chosen, we add it on top of the existing secondary sessions. Table 8 shows our results. The first column in the table shows the increasing number of secondary sessions. The second and third columns show the utility values of the secondary sessions under UPS and interweave, respectively. Comparing these two columns, we can find that the maximum number of the secondary sessions that can be supported under UPS (8) is larger than that under interweave (4). Further, for the same number of secondary sessions (from 1 to 8), the achieved utility value under UPS is higher than that under interweave.

7 Conclusions and Further Work

In this paper, we develop a policy-based network cooperation paradigm as a new dimension for spectrum sharing between the primary and secondary users. Such network cooperation can be defined as a set of policies under which different degrees of cooperation are to be achieved. The benefits of this paradigm are numerous, including improved network connectivity and spatial diversity, increased flexibility in scheduling and routing, cost savings in infrastructure needed for each individual network, among others. For the purpose of performance study, we consider a specific policy called UPS, which allows a complete cooperation between the primary and secondary networks at the node level to relay each other's traffic. We studied a problem with the goal of supporting the rate requirement of the primary network traffic while maximizing the throughput of the secondary sessions. Through rigorous mathematical modeling, problem formulation, approximation solution, and simulation results, we showed that the UPS offers significantly better throughput performance than that under the interweave paradigm.

In our future work, we will explore other policies under the policy-based network cooperation paradigm. Under a given policy, data forwarding behavior may also be affected by user requirements and performance objectives. Such user requirements and performance objectives under a particular policy are many, and each scenario would result in different data forwarding for both the primary and the secondary sessions. Clearly, there is a large landscape for further research under this new paradigm. We hope our vision and results in this paper will open the door for further research in this area.

ACKNOWLEDGMENTS

This work was supported in part by US NSF grants 1343222, 1064953, 1405747, 1443889, and ONR Grant N000141310080. The work of S. Kompella was supported in part by the ONR. Part of W. Lou's work was completed while she was serving as a Program Director at the NSF. Any opinion, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not reflect the views of the NSF.

REFERENCES

[1] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Comput. Netw.*, vol. 50, no. 13, pp. 2127–2159, Sep. 2006.

[2] I. F. Akyildiz, W.-Y. Lee, and K. R. Chowdhury, "CRAHNs: Cognitive radio ad hoc networks," *Ad-Hoc Netw.*, vol. 7, no. 5, pp. 810–836, Jul. 2009.

[3] K. R. Chowdhury and I. F. Akyildiz, "CRP: A routing protocol for cognitive radio ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 794–804, Apr. 2011.

[4] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[5] F. Gao, R. Zhang, Y.-C. Liang, and X. Wang, "Design of learning-based MIMO cognitive radio systems," *IEEE Trans. Veh. Technol.*, vol. 59, no. 4, pp. 1707–1720, May 2010.

[6] S. Geirhofer, L. Tong, and B. M. Sadler, "Cognitive radios for dynamic spectrum access—Dynamic spectrum access in the time domain: Modeling and exploiting white space," *IEEE Commun. Mag.*, vol. 45, no. 5, pp. 66–72, May 2007.

[7] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive Radios: An information theoretic perspective," *Proc. IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.

[8] Y. T. Hou, Y. Shi, and H. D. Sherali, "Spectrum sharing for multi-hop networking with cognitive radios," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 146–155, Jan. 2008.

[9] S. Hua, H. Liu, M. Wu, and S. S. Panwar, "Exploiting MIMO antennas in cooperative cognitive radio networks," in *Proc. IEEE INFOCOM*, Shanghai, China, Apr. 10–15, 2011, pp. 2714–2722.

[10] S. K. Jayaweera, M. Bkassiny, and K. A. Avery, "Asymmetric cooperative communication based spectrum leasing via auctions in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2716–2724, Aug. 2011.

[11] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Informa. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[12] J. Liu, Y. T. Hou, and H. D. Sherali, "Routing and power allocation optimization for MIMO-based ad hoc networks with dirty paper coding," in *Proc. IEEE Int. Conf. Commun.*, Beijing, China, May 19–23, 2008, pp. 2859–2864.

[13] S.-J. Kim and G. B. Giannakis, "Optimal resource allocation for MIMO ad hoc cognitive radio networks," *IEEE Trans. Inform. Theory*, vol. 57, no. 5, pp. 3117–3131, May 2011.

[14] R. Manna, R. H. Y. Louie, Y. Li, and B. Vucetic, "Cooperative spectrum sharing in cognitive radio networks with multiple antennas," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5509–5522, Nov. 2011.

[15] T. Nadkar, V. Thumar, G. Shenoy, A. Mehta, U. B. Desai, and S. N. Mechant, "A cross-layer framework for symbiotic relaying in cognitive radio networks," in *Proc. IEEE Symp. New Frontiers Dyn. Spectrum Access Netw.*, Aachen, Germany, May 3–6, 2011, pp. 498–509.

[16] President's Council of Advisors on Science and Technology (PCAST) (2012, Jul.). Report to the President—Realizing the full potential of government-held Spectrum to spur economic growth. [Online]. Available: http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast_spectrum_report_final_july_20_2012.pdf

[17] B. Radunovic and J.-Y. Le Boudec, "Rate performance objectives of multi-hop wireless networks," in *Proc. IEEE 23rd Annu. Joint Conf. INFOCOM Soc.*, Hong Kong, China, Mar.. 7–11, 2004, vol. 3, pp. 1916–1927.

[18] S. Rosloniec, *Fundamental Numerical Methods for Electrical Engineering*. Berlin, Germany: Springer, 2008.

[19] O. Simone, I. Stanojev, S. Savazzi, Y. Bar-Ness, U. Spagnolini, and R. Pickholtz, "Spectrum leasing to cooperating secondary ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 203–213, Jan. 2008.

[20] W. Su, J. D. Natyjas, and S. Batalama, "Active cooperation between primary users and cognitive radio users in cognitive ad-hoc network," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Mar. 14–19, 2010, pp. 3174–3177.

[21] A. M. Wyglinski, M. Nekovee, and Y. T. Hou, *Cognitive Radio Communications and Networks: Principles and Practice*. New York, NY, USA: Academic, 2010, Ch. 12.

[22] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[23] X. Yuan, Y. Shi, Y. T. Hou, W. Lou, and S. Kompella, "UPS: A united cooperative paradigm for primary and secondary networks," in *Proc. IEEE 10th Int. Conf. Mobile Ad-Hoc Sensor Syst.*, Hangzhou, China, Oct. 14–16, Oct. 2013, pp. 78–85.

[24] X. Yuan, C. Jiang, Y. Shi, Y. T. Hou, W. Lou, and S. Kompella, "Beyond interference avoidance: On transparent coexistence for multi-hop secondary CR networks," in *Proc. IEEE 10th Annu. Commun. Soc. Conf. Sensor, Mesh, Ad-Hoc Commun. Netw.*, New Orleans, LA, USA, Jun. 2013, pp. 398–405.

[25] R. Zhang and Y.-C. Liang, "Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 88–102, Feb. 2008.

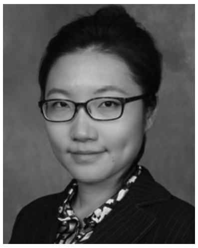
[26] J. Zhang and Q. Zhang, "Stackelberg game for utility-based cooperative radio network," in *Proc. 10th ACM Int. Symp. Mobile Ad-Hoc Netw. Comput.*, New Orleans, LA, USA, May 18–21, 2009, pp. 23–32.

[27] Y. J. Zhang and A. M.-C. So, "Optimal spectrum sharing in MIMO cognitive radio networks via semidefinite programming," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 362–373, Feb. 2011.

[28] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.

Xu Yuan (S'13-M'16) received the BS degree in computer science from the Nankai University, Tianjin, China, in 2009. Since the Fall 2010, he has been working toward the PhD degree in the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA. His current research interests include algorithm design and optimization for cognitive radio networks. He is a member of the IEEE.

Yi Shi (S'02-M'08-SM'13) received the PhD degree in computer engineering from the Virginia Tech, Blacksburg, VA, USA in 2007. He is currently a senior research scientist at Intelligent Automation, Inc., Rockville, MD, USA, and an adjunct assistant professor at Virginia Tech. His research interests include optimization and algorithm design for wireless networks. He received the IEEE INFOCOM 2008 Best Paper Award and the only Best Paper Award Runner-Up of IEEE INFOCOM 2011. He is a senior member of the IEEE.



Xiaoqi Qin (S'13) received the BS and MS degree in computer engineering from the Virginia Tech, Blacksburg, VA, USA, in 2011 and 2013, respectively. Since Fall 2013, she has been working toward the PhD degree in the Bradley Department of Electrical and Computer Engineering, Virginia Tech. Her current research interests include algorithm design and cross-layer optimization for wireless networks. She is a student member of the IEEE.

Y. Thomas Hou (F'14) received the PhD degree from NYU Polytechnic School of Engineering (formerly Polytechnic Univ.), New York, NY, USA. He is currently the Bradley distinguished professor of electrical and computer engineering, Virginia Tech, Blacksburg, VA, USA. His current research interest include developing innovative solutions to complex cross-layer optimization problems in wireless and mobile networks. He has published two graduate textbooks: *Applied Optimization Methods for Wireless Networks* (Cambridge University Press, 2014) and *Cognitive Radio Communications and Networks: Principles and Practices* (Academic Press/Elsevier, 2009). He is the steering committee chair of IEEE INFOCOM conference and a member of the IEEE Communications Society Board of Governors. He is a fellow of the IEEE.

Wenjing Lou (F'15) received the PhD degree in electrical and computer engineering from the University of Florida, Gainesville, FL, USA. She is currently a professor in the Department of Computer Science, Virginia Tech, Blacksburg, VA, USA. Her research interests include the broad area of wireless networks, with special emphases on wireless security and cross-layer network optimization. Since August 2014, she has been serving as a program director at the US National Science Foundation. She is the steering committee chair of the IEEE Conference on Communications and Network Security (CNS). She is a fellow of the IEEE.

Sastry Kompella (S'04-M'07-SM'12) received the PhD degree in computer engineering from the Virginia Tech, Blacksburg, VA, USA, in 2006. He is currently the head of wireless network theory section, Information Technology Division, U.S. Naval Research Laboratory, Washington, DC, USA. His research interests include complex problems in cross-layer optimization and scheduling in wireless and cognitive radio networks. He is a senior member of the IEEE.

Scott F. Midkiff (S'82-M'85-SM'92) is currently a professor and the vice president for information technology and chief information officer at Virginia Tech, Blacksburg, VA, USA. From 2009 to 2012, he was the department head of the Bradley Department of Electrical and Computer Engineering at Virginia Tech. From 2006 to 2009, he served as a program director at the US National Science Foundation. His research interests include wireless and ad hoc networks, network services for pervasive computing, and cyber-physical systems. He is a senior member of the IEEE.

Jeffrey H. Reed (F'05) received the PhD degree from the University of California, Davis, CA, USA in 1987. He is the Willis G. Worcester professor in the Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic and State University, Blacksburg, VA, USA. From June 2000 to June 2002, he was the director of the Mobile and Portable Radio Research Group. In 2005, he founded Wireless at Virginia Tech and served as its director until 2014. He is co-founder of Cognitive Radio Technologies, LLC, and Power Fingerprinting, Inc. His area of expertise is in software radios, cognitive radio, wireless networks, and communications signal processing. He is a distinguished lecturer for the IEEE Vehicular Technology Society. He is a fellow of the IEEE.