UPS: A United Cooperative Paradigm for Primary and Secondary Networks

Xu Yuan† Yi Shi† Y. Thomas Hou† Wenjing Lou† Sastry Kompella‡
† Virginia Polytechnic Institute and State University, USA
‡ U.S. Naval Research Laboratory, Washington, DC, USA

***Abstract*—The dominant spectrum sharing paradigm of today is the interweave paradigm. This paper advocates a new and alternative paradigm called United network of Primary and Secondary networks (UPS). UPS allows a complete cooperation between primary and secondary networks at the node level to relay each other's traffic, in addition to existing dynamic spectrum access (DSA) in time, space, and frequency domains. Such cooperation allows the primary and secondary networks to access a much richer network resources from the combined network. As a case study, we consider a problem with the goal of supporting the rate requirement of the primary network traffic while maximizing the minimum throughput of the secondary sessions. For this problem, we develop an optimization model and formulate a combinatorial optimization problem. Although this problem is in the form of mixed integer linear program (MILP), we can use CPLEX to solve it efficiently. Simulation results show that the UPS paradigm offers much better throughput performance than the interweave DSA paradigm.**

I. Introduction

The prevailing paradigm for spectrum sharing between primary and secondary networks considered in the research community is called "interweave" [4]. Under this paradigm, secondary cognitive radio (CR) nodes attempt to scavenge wireless spectrum that is not used by the primary network. This can be done by having the secondary nodes exploit transmission opportunities in time, space, and frequency domains [1], [3], [14]. Under the interweave paradigm, there is a clear separation between the primary and secondary networks, in the sense that there is no cooperation between the two networks for data forwarding.

Recently, there are some efforts on having secondary network help relay traffic for the primary network. These research were motivated by the fact that given that the secondary network is co-located with the primary network in the same geographical region, the primary network may take advantage of the secondary network, at the node level, to help forward its data. To date, efforts along this direction (see, e.g., [5], [6], [8], [9], [12], [13], [15]) have been limited to only having secondary nodes help relaying primary users' traffic. There is no consideration of the converse (i.e., primary helping the secondary), or a broader vision of a policy-base cooperation between the two networks. Such limitation is mainly due to the mindset by the current FCC rules on existing wireless services and applications. However, as user application and service requirements evolve over time, there is no reason why a primary network and a secondary network cannot cooperate on the data plane for the greater benefit of both networks.

For correspondence, please contact Prof. Tom Hou (thou@vt.edu).

In this paper, we envision United network of Primary and Secondary networks (UPS) that allows a complete cooperation between primary and secondary networks to relay each other's traffic. Although such a paradigm may be ahead of its time in practice, there is no reason why one should not investigate its capability and potential from research perspective. The UPS paradigm allows to pool together the resources from both the primary and secondary networks so that users in each network can access a much richer network resources from the combined network. There are many potential benefits of UPS, such as much improved network topology, opportunity of better power control, more flexibility in link layer scheduling and network layer routing, and a much richer set of service offerings for users in the primary and secondary networks. Note that the vision of UPS can still be configured to preserve the special "privilege" or priority requirements that are previously offered to the primary network. Such priority or guaranteed services can easily be supported by implementing appropriate administrative policies in the combined network.

As a case study of the UPS paradigm, we consider a problem with the goal of supporting the rate requirements of the primary sessions while maximizing the minimum throughput of the secondary sessions. Since the primary network is the owner of the spectrum, we may offer certain priority or guaranteed service to primary traffic over the secondary traffic in the combined network. On the other hand, we may not offer any guarantee for secondary network traffic, which will be supported on a best-effort basis based on any remaining resources in the combined network. A number of technical challenges must be addressed in this problem, such as how to provide guaranteed service for primary traffic while supporting as much as 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 linear program (MILP), we can use commercial software (e.g., CPLEX) to solve it efficiently. Through simulation results, we demonstrate that UPS paradigm can indeed offer much better throughput performance than the existing interweave paradigm.

The remainder of this paper is organized as follows. In

Section II, we describe the state-of-the-art of spectrum sharing paradigms. We also make a case for the UPS paradigm for broader applications and discuss its benefits. Under the UPS paradigm, we consider a case study in Section III, and Section IV presents an optimization model for the case study. Section V presents simulation result and Section VI concludes this paper.

II. The Case of UPS

A. State of the Art

Prevailing Paradigm. For resource sharing between primary and secondary networks, the prevailing paradigm considered in the research community is called "interweave" [4]. Under this paradigm, the primary network "owns" the spectrum and uses it without any concern of the secondary network. A secondary network attempts to scavenge wireless spectrum that is not used by the primary network for its own communication. This can be done by having the secondary network exploit transmission opportunities in time, space, and frequency domains [1], [3], [14].

The essence of the interweave paradigm is opportunities transmission. That is, the secondary network is only allowed to exploit remaining radio resources in the background of primary network activities. It is not allowed to interfere with (or be noticeable to) the primary network. The rationale behind this paradigm is that the primary network may be the traditional wireless network who owns its spectrum, while the secondary network is likely to be highly intelligent CRs and can only use the same spectrum when they are not interfering with the primary network. So there shall be no additional requirements on the primary network for spectrum sharing and all burden for opportunities transmission shall rest upon the secondary network. Under the interweave paradigm, there is no "mutual" interaction between the primary and secondary networks in data forwarding, despite that the secondary network pro-actively observes primary network's activities and exploits any remaining radio resources. That is, there is no node-level resource sharing between the two networks.

Cooperative Relaying. Recently, there are some research efforts on having the secondary network help relay traffic for the primary network. This research was motivated by the fact that given that the secondary network is co-located with the primary network in the same geographical region, the primary network may take advantage of the secondary network, at the node level, to help its data transport. In [12], 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 in relaying its data. In [15], 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 [13], Su *et al.* proposed to have the primary network lease its spectrum in the frequency domain to the secondary network to relay it data in order to maximize the primary users' energy saving and the secondary users' data rates. In [6], Jayaweera *et al.* proposed a new way of encouraging primary users to lease their spectrum by having the secondary users to place bids on the amount of power they are willing to expend for relaying primary users' traffic. In [5], 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 [8], Manna *et al.* considered the three-node model in [7]. The relay node is 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 [9], Nadkar *et al.* considered how to offer incentive (in terms of time and frequency) to a secondary network to help transmit primary user's traffic. They studied a cross-layer optimization problem that maximizes the transmission opportunities for the 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 have secondary nodes help primary nodes in relaying primary users' traffic. This is, however, only half of the story. In this paper, we advocate a much greater cooperation between the two networks.

B. UPS: An Overview

We envision United network of Primary and Secondary networks (or "UPS" in short) that allows complete cooperation between the two networks on the data plane in terms of relaying *each other*'s traffic. Unlike previous efforts on node-level cooperative relaying, which was limited to only allowing secondary nodes to help relay primary users' traffic, UPS allows primary nodes to help relay traffic for the secondary network. More important, the UPS paradigm allows to pool all the resources from the primary and secondary networks together and allows users in each network to access a much richer network resources from the combined network. Figure 1 illustrates the concept of UPS, where the two networks united together to form one combined network. Although the two networks are combined into one at the physical level, priority or service guarantee to the primary network traffic can still be offered at the transport level, by implementing certain traffic engineering policies.

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

- **Topology.** Comparing to primary or secondary network in isolation, 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.

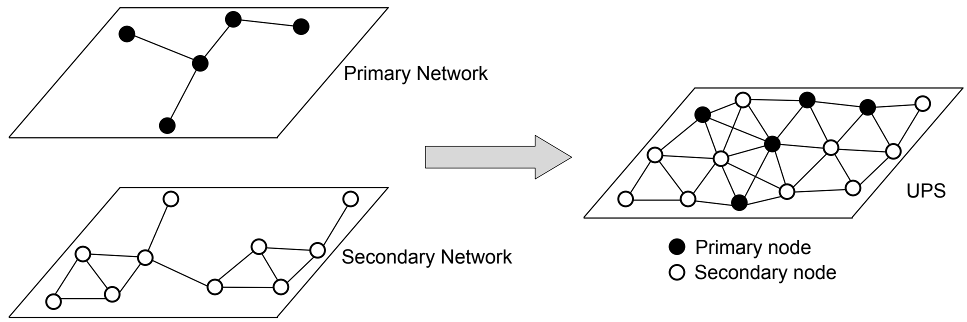

Fig. 1. Network topologies under the interweave and the UPS paradigms.

- **Link Layer.** The improved physical topology allows more opportunities (as well as challenges) at the link layer for medium 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 [5], [6], [8], [9], [12], [13], [15]. Although the two networks are combined, the service 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.

III. Problem Scope

In the rest of this paper, we consider a case study under the UPS paradigm. Suppose there is a set of sessions in the primary network, with each session having 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 hard rate requirement as a primary session. Instead, each secondary session will be supported on a best-effort basis and will transmit as much as the network resources allow. The goal is to have the combined network to support the rate requirements of the primary sessions while maximizing the minimum throughput of the secondary sessions.

For this problem, there are a number of technical challenges that one must address under the UPS paradigm.

- **Guaranteed service for primary traffic.** In the above problem, each primary session has hard rate requirement and 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 provide guaranteed service for primary traffic and to maximize the minimum rate of the secondary sessions, relay node selection along a route (for either a primary or secondary session) is a key problem.
- **Scheduling.** To achieve both objectives in the primary and secondary networks, scheduling in each time slot is not a trivial problem. In particular, in addition to addressing traditional self-interference (half-duplex) and mutual interference problems, primary network must be cooperative so as to help the secondary sessions to achieve their optimization objective in the combined network. Such behavior from the primary network is unique under the UPS paradigm and has never been explored before.

IV. Modeling and Formulation

In this section, we develop a mathematical model for the UPS paradigm. With this model, we consider a throughput maximization problem for the secondary users.

A. Mathematical Modeling

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) 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 $\mathcal{I}_j$ as the set of nodes (including both primary and secondary nodes) located within a 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.

Guaranteed Service for the Primary Sessions. Under the UPS paradigm, the primary sessions consider the combined network $\mathcal{N}$ as their usable 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 paradigm, the primary sessions have priority in access the combined network resources (in the form of guaranteed services). While the primary sessions are supported, the secondary sessions may use as much as the remaining resources of the combined network. How the primary and secondary sessions interact in the combined network is a key part of our 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

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

- 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, we will only need to optimize the flow paths in (1)–(3), while we need to both optimize the routes and maximize the objective $r(m)$ in (4)–(6).

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 $1 \leq t \leq T$, to indicate whether node i transmits data to node j. That is, if node i transmits data to node j, $x_{ij}[t] = 1$; otherwise, $x_{ij}[t] = 0$.

Since 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

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

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

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

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

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

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

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. 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:

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

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

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

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

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

Link Capacity Constraints. For each link (i, j), denote the link capacity as C_{ij}. Since the aggregate flow rate from the primary and secondary sessions on each link (i, j) cannot exceed the average link rate (over T time slots), we have

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

B. 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 are assumed to be elastic. To ensure fairness among the sessions, we set our objective function to maximize the minimum session rate among all secondary sessions. We define $r_{\min}$ as this minimum rate. Then we have:

$$r_{\min} \leq r(m) \qquad (m \in \mathcal{L}). \tag{14}$$

The optimization problem can be written as follows:

OPT
max $r_{\min}$
s.t. Throughput for secondary sessions: (14),
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, $\hat{f}_{ij}(l), f_{ij}(m), r(m)$ and $r_{\min}$ are continuous variables. This problem is in the form of *mixed integer linear program* (MILP). Although the theoretical worst-case complexity to a general MILP problem is exponential [2], [11], we found that OPT can be solved by CPLEX efficiently, due to fact that all integer variables $x_{ij}[t]$ are binary.

V. Simulation Results

In this section, we present numerical results to demonstrate the capabilities and advantages of the UPS paradigm. The goal of this section is twofold. First, we show that the UPS paradigm offer much better performance for both 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 paradigm.

A. 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 transmission power of each node $i \in \mathcal{N}$ is set to $Q_i = 1$. We assume the transmission range and interference range at all nodes are 30 and 50, respectively. The number of time slots in a frame is $T = 10$. A link's capacity is calculated by $C_{ij} = B\log_2(1+\frac{Q_i d_{ij}^{-4}}{N_0})$, where d_{ij} is the distance between nodes i and j, and N_0 is the ambient Gaussian noise density. We assume $N_0 = 10^{-6}$.

B. Results

We consider a 30-node network, with 15 primary nodes and 15 secondary nodes randomly deployed in a 100×100 area (see Fig. 2). 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 I. 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 paradigm and the interweave paradigm, respectively, and (ii) the objective value of secondary session rate in our optimization problem under the two paradigms. The optimization problem for maximizing the minimum secondary session rate under interweave paradigm can be formulated following a similar token to OPT and is given in the appendix.

Table II summarizes the results of this study. The second column represents increasing rate requirements for the primary sessions (i.e., $\hat{R}(1)$ and $\hat{R}(2)$). For ease of explanation, we break this table into five regions, with each region representing an operating behavior for comparison under the two paradigms. The third and fourth columns show the performance under the UPS paradigm. Specifically, the third column shows whether the rate requirements of the two primary sessions are feasible in the primary network (abbreviated as "PN" in the table); the fourth column shows the maximized minimum data rate between the two secondary sessions (abbreviated as "SS" in the table) with 0 indicating zero rates for the secondary sessions 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 paradigm, respectively.

Region 1 This region represents the scenario where the rate requirements of the primary sessions can be supported under both paradigms *and* the rates of the secondary sessions are positive. Comparing columns four and six, we can find that the secondary sessions always achieve higher performance under the UPS paradigm than that under the interweave paradigm.

TABLE I

THE SOURCE AND DESTINATION NODES FOR EACH SESSIONS IN THE 30-NODE NETWORK.

Session	Source	Destination
Primary session 1	P_{10}	P_7
Primary session 2	P_{15}	P_1
Secondary session 1	S_6	S_{15}
Secondary session 2	S_{12}	S_3

TABLE II

PERFORMANCE COMPARISON BETWEEN THE UPS 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 rate	Feasible in PN	SS rate
Region 1	(0, 0)	Yes	6.0073	Yes	4.8277
	(0.2, 0.3)	Yes	5.5229	Yes	3.4457
	(0.4, 0.6)	Yes	5.5229	Yes	3.4457
	(0.6, 0.9)	Yes	5.5.5229	Yes	3.4457
	(0.8, 1.2)	Yes	5.5229	Yes	3.4457
	(1.0, 1.5)	Yes	4.6263	Yes	3.378
	(1.2, 1.8)	Yes	4.6263	Yes	3.378
	(1.4, 2.1)	Yes	4.6263	Yes	2.4923
	(1.6, 2.4)	Yes	4.6263	Yes	1.7958
	(1.8, 2.7)	Yes	4.0019	Yes	1.2461
	(2.0, 3.0)	Yes	4.0019	Yes	1.2461
Region 2	(2.2, 3.3)	Yes	2.9815	Yes	0
	(2.4, 3.6)	Yes	2.7795	Yes	0
	(2.6, 3.9)	Yes	2.7795	Yes	0
	(2.8, 4.2)	Yes	2.7795	Yes	0
	(3.0, 4.5)	Yes	2.7795	Yes	0
	(3.2, 4.8)	Yes	2.7795	Yes	0
Region 3	(3.4, 5.1)	Yes	2.7795	No	N/A
	(3.6, 5.4)	Yes	1.4907	No	N/A
Region 4	(3.8, 5.7)	Yes	0	No	N/A
	(4.0, 6.0)	Yes	0	No	N/A
	(4.2, 6.3)	Yes	0	No	N/A
	(4.4, 6.6)	Yes	0	No	N/A
	(4.6, 6.9)	Yes	0	No	N/A
	(4.8, 7.2)	Yes	0	No	N/A
	(5.0, 7.5)	Yes	0	No	N/A
	(5.2, 7.8)	Yes	0	No	N/A
	(5.4, 8.1)	Yes	0	No	N/A
	(5.6, 8.4)	Yes	0	No	N/A
	(5.8, 8.7)	Yes	0	No	N/A
	(6.0, 9.0)	Yes	0	No	N/A
	(6.2, 9.3)	Yes	0	No	N/A
	(6.4, 9.6)	Yes	0	No	N/A
	(6.6, 9.9)	Yes	0	No	N/A
Region 5	(6.8, 10.2)	No	N/A	No	N/A

This confirms our expectation that the UPS paradigm can offer higher throughput for the secondary sessions.

As an example, consider the case when the two primary sessions have rate requirements (2.0, 3.0). The objective value achieved for the secondary sessions under the UPS and the interweave paradigms are 4.0019 and 1.2461, respectively. Under UPS paradigm, the flow routing and scheduling for the primary and secondary sessions are shown in Fig. 2. The number in the box on each link represents the active time slots for this link. Note that primary nodes P_4, P_5, P_6, P_{11}, and P_{13} are helping relay secondary sessions' data while secondary nodes S_2, S_8 and S_{13} are helping relay the primary sessions' data. Under the interweave paradigm, the flow routing and scheduling for primary network is shown in Fig. 3, which can

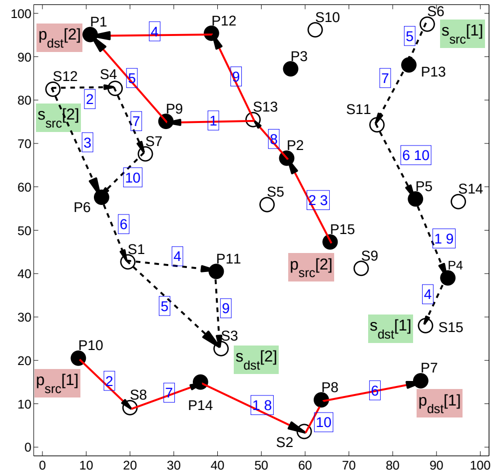

Fig. 2. A Region 1 example that shows the flow routing topologies for the primary and secondary sessions in the UPS paradigm, where solid line segments are for the primary sessions while dashed line segments are for the secondary sessions.

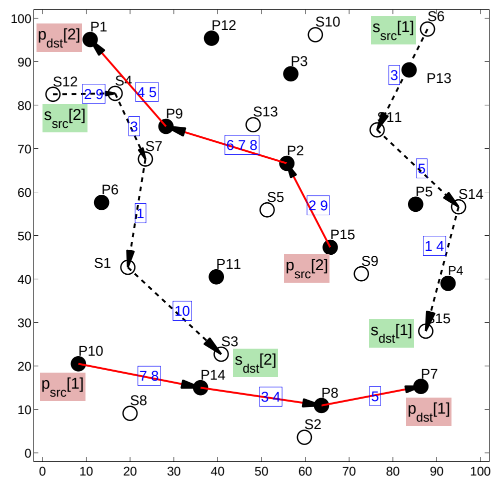

Fig. 3. A Region 1 example that shows the flow routing and scheduling for each primary session and secondary session under the interweave paradigm.

satisfy rate requirements for primary sessions. According to the time slots used by the primary network, the secondary network calculate the available time slot at each node and use them to maximize their minimum data rate among all sessions. The flow routing and scheduling for the secondary sessions under interweave paradigm are also shown in Fig. 3. 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 paradigms while the secondary sessions can only be supported under the UPS paradigm but not under the interweave paradigm (with $r_{\min} = 0$). This region shows 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 (3.2, 4.8). The minimum

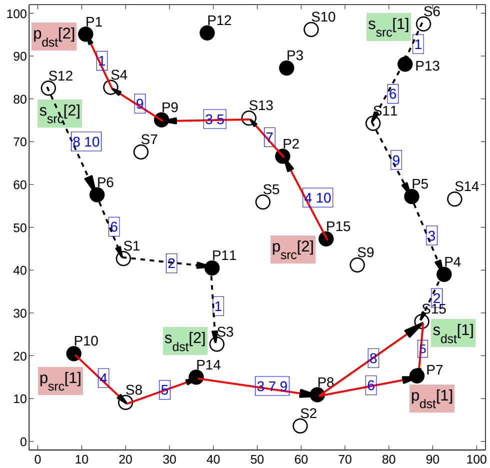

Fig. 4. A Region 2 example that shows the flow routing topologies and scheduling for the primary and secondary sessions in the UPS paradigm.

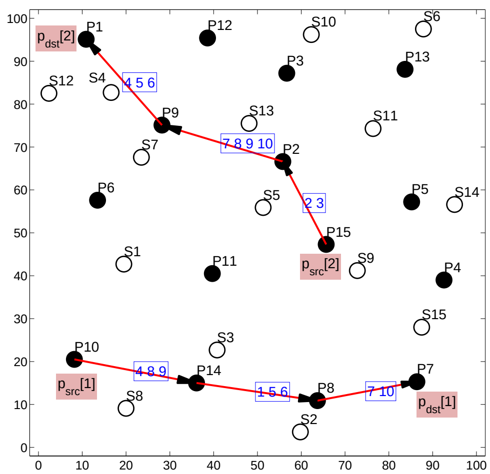

Fig. 5. A Region 2 example that shows the flow routing topologies and scheduling for primary network in the interweave paradigm.

data rate achieved for the secondary sessions under the UPS paradigm is 2.7995. Under the UPS paradigm, the flow routing and scheduling for the primary and secondary sessions are shown in Fig. 4. Note that primary nodes P_4, P_5, P_6, P_{11} and P_{13} are helping to relay secondary sessions data while secondary nodes S_4, S_8, S_{13} and S_{15} are helping to relay the primary sessions data. Under the interweave paradigm, the flow routing and scheduling for primary network are shown in Fig. 5. According to the time slots used by the primary network, the secondary network calculate the available time slot at each node. However, the remaining time slots are not enough to support the secondary sessions, resulting in at least one of the sessions with zero rate.

Region 3 This region represents the scenario where the rate requirements of the primary sessions can be supported under the UPS paradigm but not so under the interweave paradigm. For secondary sessions, there is still remaining resource to support them under the UPS paradigm. For fairness in comparison, we do not consider the achieved rate of the

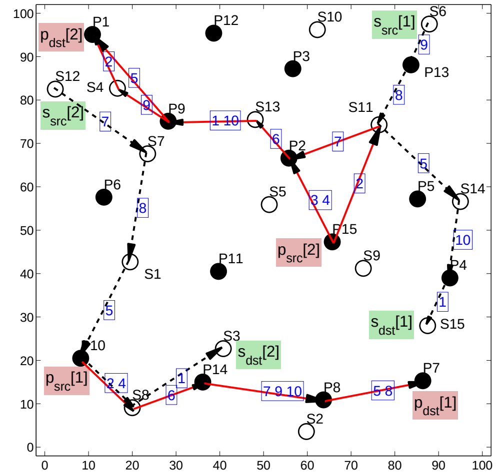

Fig. 6. A Region 3 example that shows the flow routing topologies and scheduling for primary and secondary sessions under UPS paradigm.

secondary sessions under the interweave paradigm (marked as "N/A"). The region shows the definitive advantage of using a combined network to support the primary sessions over an independent primary network.

As an example, we consider the case when the two primary sessions have rate requirements (3.6, 5.4). The objective value achieved by secondary sessions is 1.69878 under the UPS paradigm. Under UPS paradigm, the flow routing and scheduling for primary and secondary sessions are shown in Fig. 6. Note that primary nodes P_4, P_{10} and P_{13} are helping to relay secondary sessions data while secondary nodes S_4, S_8, S_{13}, S_{15} are helping to relay the primary data.

Region 4 This region represents the scenario where the rate requirement of the primary sessions can be satisfied under the UPS paradigm but not so under the interweave paradigm. The secondary sessions can no longer be supported under the UPS paradigm (with $r_{min} = 0$). For fairness in comparison, we do not consider the achieved data rate 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, Region 4 shows the advantage of using a combined network to support the primary sessions over an independent primary network.

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

VI. Conclusions

In this paper, we presented the UPS paradigm as a new dimension for further spectrum sharing paradigms. UPS allows complete cooperation between the primary and secondary networks at the node level and on the data plane to help relay each other's traffic. This is a major step forward beyond existing interweave DSA paradigm or unilateral cooperation only from the secondary network. There are many benefits associated with the UPS paradigm, which span from physical topology, multiple layers of the protocol stack, and upper layer performance. Although the primary and secondary networks are combined into one network at the physical level, various service differentiation policies (including those policies of today) on the primary and secondary traffic can still be supported by implementing appropriate traffic engineering policies. To illustrate this point, we considered a case study with the goal of supporting the rate requirements of the primary sessions while maximizing the minimum rate among the secondary sessions. Through a systematic mathematical model, problem formulation, solution development, and simulation study, we showed that the UPS paradigm offers a definitive advantage over the existing interweave DSA paradigm. The UPS paradigm may be ahead of its time per today's FCC policies. But our study offers a compelling reason for further exploring this new paradigm, which could well evolve into a viable approach for future spectrum sharing in the real world.

ACKNOWLEDGMENTS

This research was supported in part by NSF Grants 0831865, 1343222, 1064953, 1156318, 1156311, and ONR Grant N000141310080. The work of S. Kompella was supported by the ONR.

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," *Elsevier Computer Networks,* vol. 50, no. 13, pp. 2127–2159, September 2006.

[2] M.R. Garey and D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness,* W.H. Freeman and Company, New York, 1979.

[3] S. Geirhofer, L. Tong, and B.M. Sadler, "Dynamic spectrum access in the time domain: Modeling and exploiting white space," *IEEE Communications Magazine,* vol. 45, no. 5, pp. 66–72, May 2007.

[4] A. Goldsmith, S.A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: an information theoretic perspective," *Proceedings of the IEEE,* vol. 97, no. 5, pp. 894–914, May 2009.

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

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

[7] J.N. Laneman, D.N.C. Tse, and G.W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. on Information Theory,* vol. 50, issue 12, pp. 3062–3080, Dec. 2004.

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

[9] 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 DySPAN,* pp. 498–509, Aachen, Germany, May 2011.

[10] B. Radunovic and J.-Y. Le Boudec, "Rate performance objectives of multi-hop wireless networks," in *Proc. IEEE INFOCOM,* pp. 1916–1927, Hong Kong, China, March 7–11, 2004.

[11] A. Schrijver, *Theory of Linear and Integer Programming,* Chapter 18, Wiley Interscience, New York, NY, 1998.

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

[13] 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 ICASSP,* pp. 3174–3177, March 2010.

[14] A.M. Wyglinski, M. Nekovee, and Y.T. Hou, *Cognitive Radio Communications and Networks: Principles and Practice,* Chapter 12, Academic Press/Elsevier, 2010.

[15] J. Zhang and Q. Zhang, "Stackelberg game for utility-based cooperative radio network," in *Proc. ACM MobiHoc,* pp. 23–32, New Orleans, Louisiana, USA, May 18–21, 2009.

APPENDIX – INTERWEAVE PARADIGM

Under the interweave paradigm, a primary network operates independently without any concern of the secondary network's activity. For the secondary network, it needs to sense the primary network environment and identify the available time slots for transmission based on interference relationship with nearby primary nodes. For each secondary node, it may have different available time slots set when it works as a transmitter or receiver. We denote $\mathcal{Z}_{ij}$ as the set of common available time slots for secondary transmitter $i \in \mathcal{N}_{\rm S}$ and secondary receiver $j \in \mathcal{N}_{\rm S}$. Then, the network throughput optimization problem for secondary network under interweave paradigm can be formulated as following:

OPT-Interweave

$$
\begin{aligned}
\max \quad & r_{\min} \\
\text{s.t.} \quad & r_{\min} \le r(m) \qquad (m \in \mathcal{L}), \\
& \textstyle\sum_{j \in \mathcal{A}_i} f_{ij}(m) = r(m) \qquad (m \in \mathcal{L}, i = s(m)), \\
& \textstyle\sum_{j \in \mathcal{A}_i}^{j \ne s(m)} f_{ij}(m) = \sum_{k \in \mathcal{A}_i}^{k \ne d(m)} f_{ki}(m) \\
& \qquad (m \in \mathcal{L}, i \in \mathcal{N}_{\rm S}, i \ne s(m), d(m)), \\
& \textstyle\sum_{j \in \mathcal{A}_i} x_{ij}[t] + \sum_{k \in \mathcal{A}_i} x_{ki}[t] \le 1 \\
& \qquad (i \in \mathcal{N}_{\rm S}, t \in \mathcal{Z}_{ij}), \\
& \textstyle\sum_{i \in \mathcal{A}_j} x_{ij}[t] + \sum_{k \in \mathcal{A}_p} x_{pk}[t] \le 1 \\
& \qquad (p \in \mathcal{D}_j, j \in \mathcal{N}_{\rm S}, j \ne k, t \in \mathcal{Z}_{ij}), \\
& \textstyle\sum_{m \in \mathcal{L}}^{j \ne s(m), i \ne d(m)} f_{ij}(m) \le \frac{1}{T} \sum_{t \in \mathcal{Z}_{ij}} C_{ij} \cdot x_{ij}[t] \\
& \qquad (i \in \mathcal{N}_{\rm S}, j \in \mathcal{A}_i),
\end{aligned}
$$

where $\mathcal{A}_i$ represents the set of secondary nodes that are located within node i's transmission range ($i \in \mathcal{N}_{\rm S}$), and have common available time slots with node i. $\mathcal{D}_i$ represents the set of secondary nodes that are located within node i's interference range ($i \in \mathcal{N}_{\rm S}$), and have common available time slots with node i. The first and second constraints are flow balance constraints. The third constraint is for self interference. The four constraint is for mutual interference and the last constraint is for link capacity.

OPT-Interweave is in the form of MILP, which can be solved by CPLEX efficiently, due to fact that all integer variables $x_{ij}[t]$ are binary.