Bandwidth Partition and Allocation for Efficient Spectrum Utilization in Cognitive Communications

Song Huang, Di Yuan, and Anthony Ephremides

Abstract: **Conventional cognitive communications rely heavily on the smartness of secondary (unlicensed) users (SUs) to achieve high spectrum utilization, which involves the optimization of the SUs' policies and behaviors for dynamic spectrum access, power allocation among multiple channels, etc. Due to the inherent randomness of the primary users' (PUs') transmission, those efforts inevitably increase the implementation complexity and sensing overheads of the SUs, and in turn lower the spectrum utilization efficiency. In this paper, we try to change the focus from SU to PU. A cooperative traffic allocation strategy for PU, together with the non-uniform bandwidth partition, is employed to regularize the PU's resource occupancy pattern without compromising its performance, and to maximize the spare bandwidth for the SU at the same time. We first study the capacity based optimization problem (COP) together with the fully polynomial time approximation scheme (FPTAS) for an approximation guarantee of the global optimum. Then we analyze the subcarrier based optimization problem as the surrogate problem of COP, which can be solved by a greedy algorithm exactly. Both the theoretical analysis and the numerical simulations demonstrate the effectiveness of those methods to achieve the performance that almost identical to that of the global optimum solution.**

Index Terms: **Cognitive communication, FPTAS, non-uniform bandwidth partition.**

I. INTRODUCTION AND MOTIVATION

WITH the proliferation of wireless services, most of the available spectrum has fully been allocated, which results in the spectrum scarcity problem. However, a large portion of the licensed spectrum experiences low utilization [1]. As the spectrum scarcity is mainly caused by the inefficient and inflexible spectrum allocation policies, dynamic spectrum access (DSA) [2] has been proposed as an alternative policy to improve the spectrum efficiency. Cognitive radio [3]–[5], as one of the enabling technologies of DSA, allows unlicensed users to communicate using the licensed spectrum dynamically. Under the basic model of cognitive radio networks, secondary users (SUs) employ white spaces that are not used by the primary users (PUs) under the condition of not interfering with active PUs. Most existing literature concentrates on the optimization of SUs' policies and behaviors for dynamic spectrum sensing and access, power allocation among multiple channels, or bandwidth allocation to multiple SUs [6]–[9]. Due to the inherent randomness of the PU transmission, those efforts inevitably increase the implementation complexity and sensing overheads of the SUs, and in turn lower the spectrum utilization efficiency.

In this paper, we try to improve the spectrum utilization by regularizing the resource occupancy pattern of PU without compromising its performance. To this end, the overall bandwidth is first partitioned into subcarriers, which are then grouped into channels. The PU traffic allocation is based on multiple channels, and is designed to leave as much bandwidth resource to SUs after satisfying the PU's bandwidth demand. During this process, in order to minimize the capacity loss due to spectrum fragmentation, a non-uniform grouping scheme of the subcarriers [10] is employed. As a result of bandwidth partitioning and subcarrier grouping, spectrum sensing applies to the channels instead of all the individual subcarriers. Such a design inherently reduces the sensing overheads of SUs.

From the conventional viewpoint of cognitive radio, there is no restriction on resource allocation for a PU. However, that does not prohibit interaction and cooperation between PU and SU, with the aim of improving the utilization efficiency of the overall spectrum. Existing literature on cooperative cognitive radio networks has studied various schemes of spectrum subleasing and auctions [11]–[13]. The PU has privileges in achieving its traffic requirements and thus is more willing to sublease its power and sub-channels to obtain additional revenue rather than enhancing the performance. In return, the SU helps to relay the PU's packets.

Similarly, our proposal takes an approach by regulating the PU's resource allocation pattern (without denying its bandwidth demand) and thereby increasing the resource available to SU. In [10], we have demonstrated that under the non-uniform scheme, the average amount of sensing was reduced and the capacity loss due to spectrum fragmentation during the allocation of the PU traffic was far less than that of the uniform partition schemes. However, that discussion mainly focused on the non-fading situation. In this paper, we give a detailed discussion upon the non-uniform scheme under frequency selective fading.

Spectrum occupancy is usually defined as the probability that a measured signal is above a predefined power threshold within a certain bandwidth and a specific time span [14]. The occu-

Manuscript received August 2, 2018; approved for publication by Yousefi'zadeh Homayoun, Division II Editor, April 19, 2019.

The work of S. Huang was supported by China Scholar Council (CSC) grant No. 201206155005 and Natural Science Foundation of Guangdong Province (CHINA) grant No. S2011010004773. The work of D. Yuan has been carried out within European FP7 Marie Curie IOF project 329313. The work of A. Ephremides was supported by NSF grant No. CCF1420651 and ONR grant No. N000141410107.

S. Huang is with the School of Computer Science and Engineering, South China University of Technology (SCUT); and Communication & Computer Network Lab of GD, GZ, China 510641, email: crshuang@scut.edu.cn.

D. Yuan is with the Department of Science and Technology, Linköping University, Sweden, email: diyua@itn.liu.se.

A. Ephremides is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, 20742, USA, email: etony@umd.edu.

Digital Object Identifier: 10.1109/JCN.2019.000031

1229-2370/19/$10.00 © 2019 KICS

pancy power threshold can be set at the measured noise floor plus some margin to account for the variations of the momentary noise power. This margin must be large enough to reduce the false alarm rate due to the temporal noise variation which could result in overestimating the spectral occupancy. On the other hand, this margin cannot be too large, otherwise the temporal signal variation due to the fading effects can cause missed detections which in turn results in underestimating the spectral occupancy.

The main contributions of this paper are as follows:

1) By assuming the channel state information (CSI) of the SU under selective fading being known to the PU, we model the problem of maximizing the throughput of the SU as the *Capacity-based Optimization Problem* (COP) and prove that COP is NP-hard.

2) A fully polynomial time approximation scheme (FPTAS) is proposed for COP, which provides a guaranteed approximation to the global optimal solution to COP. By changing the approximation parameter, we can get any desired approximation accuracy.

3) For the case in which CSI of the SU is unavailable to the PU, we further propose the *Subcarrier-based Optimization Problem* (SOP) as the surrogate problem of COP, with the aim of maximizing the total number of spare subcarriers for the SU. We prove that SOP can be exactly solved by a greedy algorithm with a polynomial-time complexity.

The remainder of this paper is organized as follows. In Section II, related work is discussed. In Section III, we introduce the system model, and define the effective capacity of a channel. In Section IV, COP is presented and is proved to be NP-hard. The FPTAS algorithm is proposed to approximate the global optimal solution. In Section V, we propose SOP as the surrogate problem of COP. A greedy algorithm for SOP is given and its global optimality is proved. In Section VI, we present the simulated performance of several approaches, including the optimal solution to COP, the FPTAS to COP, and the greedy algorithm to SOP. For the sake of comparison, the outcome of the uniform scheme is given as well. In Section VII, we summarize our conclusions.

II. RELATED WORK

A cognitive radio (CR) is a concept that was first defined in [2] as a radio device that can adapt its transmitter parameters to the operating environment. It is based on the concept of software defined radio [15] that can alter parameters such as frequency band, transmission power, and modulation scheme through changes in software. SUs can detect and utilize white spaces (also known as spectrum holes [16]) that are not fully used by PUs but must avoid interfering with active PUs [17].

There are three main cognitive radio network paradigms [4], namely *underlay*, *overlay*, and *interweave*. The underlay paradigm mandates that concurrent noncognitive and cognitive transmissions may occur only if the interference generated by the SU devices at the PU receivers is below some acceptable threshold, which is also referred to as the *interference temperature* [1]. Since the interference constraints in underlay systems are typically quite restrictive, this usually limits the SUs to short range communications.

The second paradigm, namely the overlay model, requires that the SU transmitter has the knowledge of codebook and messages of the PU. Under that model, the knowledge of the PU's message and/or its codebook need to be exploited to either cancel or mitigate the interference seen at receivers of the SU and the PU. Therefore, it is assumed the PU's message is known at the SU transmitter when the PU begins its transmission. However, that assumption is usually impractical for an initial transmission.

In this paper, we focus on the interweave model, which is based on the idea of *opportunistic spectrum access* (OSA) and is also referred to as *spectrum overlay* in [18]. By following the OSA strategy, the SU carries out spectrum sensing to detect white spaces, and reconfigures its transmission parameters (e.g., carrier frequency, bandwidth, and modulation scheme) to operate in the identified spectrum band. In addition, the SU needs to keep monitoring the spectrum on which it operates and quickly vacates the band whenever the PU becomes active.

To enable the SU to efficiently fill the spectral holes left by the PU without causing unacceptable interference, an OFDM-like reconfigurable subcarrier structure is usually employed. However, the adoption of OFDM structure in cognitive communication inevitably leads to dozens or even hundreds of subcarriers, which could significantly increase the sensing overheads of the SU. To address that problem, subcarrier grouping [19], [20] can be employed, which bundles the subcarriers into groups and manages each subcarrier group as a channel in traffic allocation. To acquire the occupancy status of a channel, only a representative subcarrier of the channel needs to be sensed. Such a channel-based sensing strategy can significantly reduce the amount of sensing required.

The subcarrier grouping can be either uniform or non-uniform. Under the uniform grouping scheme, all channels have equal number of subcarriers. As the PU's data rate is not always an integer multiple of the channel capacity, there is a gap between the total channel capacity offered to the PU and the actual spectrum it actually requires. That gap could inevitably reduce the capacity available to the SU, and will lower the spectrum utilization. Here we refer to the gap as *the capacity loss due to fragmentation*. A larger group size creates less number of channels, which enables the less amount of sensing. However, it may lead to more capacity loss due to fragmentation. In contrast, a smaller group size can reduce the capacity loss due to fragmentation, but it creates more channels, which unavoidably results in more sensing overheads.

To address that conflict, we have proposed a non-uniform bandwidth partition and traffic allocation scheme in [10], in which the numbers of subcarriers within the channels form a geometric sequence. Such a grouping scheme can achieve relatively low sensing overheads and diminishing capacity loss due to fragmentation simultaneously. The scheme we proposed has neither restrictions nor influence on the PU's traffic profile, such as power of transmission, preambles, midambles and pilot patterns. Thus most existing spectrum sensing methods [21]–[25] can be adopted without altering their implementations or their properties. However, our discussion in [10] about the non-uniform scheme mainly focused on the non-fading situation. In

this paper, we concentrate on how to maximize the SU throughput of the non-uniform scheme under fading.

In [26], we have proved the NP-hardness of COP, and have presented the greedy algorithm for SOP. In the current paper, the novelties consist of the following aspects: 1) The FPTAS is derived for COP as an approximation solution, and is modeled as a knapsack problem with non-integer weight constraints. 2) As conventional dynamic programming methods cannot be employed directly, we propose a novel dynamic programming algorithm for the FPTAS. 3) Additional simulations have been made to illustrate the performance differences between the FPTAS and the greedy algorithm.

Under frequency selective fading, the effective capacity of channels varies across different bands and time slots. To maximize the spectrum utilization, we have to allocate spectrum to the PU adaptively according to its current traffic volume and the instant effective capacity of the channels, which is essentially a process of spectrum adaptation. Although the spectrum adaptation approaches have been broadly adopted in wireless communications in varies forms, such as fine-grained channel access [27], flexible channelization [28], spectrum slicing [29], channel assembling and fragmentation [30], [31], etc., there are significant differences between the spectrum adaptation approach we employed and those of existing literature, which are listed as follows:

First, our spectrum adaptation is designed for the interweave paradigm, in which the PU and the SU share no channel. The SU keeps sensing and monitoring the channels and vacates any of them that will be used by the PU. Within an idle time-spectrum block, the SU can maximize its transmission power and data rates as long as its transmission does not interference the adjacent channels. In contrast, the channel assembling and fragmentation of [31] aim to distribute the PU traffic flow evenly across all available channels. That strategy actually forms an underlay paradigm, and unavoidably limits the SU's communication distance and data rate.

Secondly, our spectrum adaptation scheme is based on the cooperation between the PU and the SU. The PU optimizes its bandwidth occupancy to minimize the capacity loss due to fragmentation and to leave as much idle bandwidth to the SU as possible. Such a cooperative strategy for traffic allocation can effectively lower the sensing overheads of the SU. In contrast, conventional spectrum adaptation requires no cooperation from the PU. It is all up to the SU to perform spectrum sensing and monitoring. The randomly generated spectrum holes, which varies in size and position, inevitably make the design and implementation of the SU complex [17].

The performance of cognitive radio networks highly depends upon the activity of primary radio users. By modeling PU's activity, SUs can predict the future state of PUs by learning from the history of their spectrum utilization, and then assign best available spectrum bands for their communication. The PU activity models employed in CRN can be classified into several categories: Markov process, queuing theory, time series, and ON/OFF periods, etc [32]–[35]. In prediction-based spectrum sensing, a CR user can skip the sensing duty on some channels that are predicted to be busy, thus reducing the sensing time and energy consumption.

As for the relationship between our scheme and PU activity models, the key points are summarized as follows: First, our scheme is based on spectrum occupancy regulation. The modeling of PU activity is a method for spectrum occupancy prediction. They are independent to each other and can be used cooperatively in practice. On one hand, our scheme aims to regulate the PU traffic allocation. That will make the spectrum occupancy become more concentrated by reducing the number of partially occupied channels, and will eventually leave more completely fully idle channels to SUs. The regulation of spectrum occupancy can ease the adoption of PU activity models. On the other hand, it is also feasible to use PU activity models to improve the performance of our scheme. At least, the correlation among the PU traffic sequences can be used to decrease the number of sensing operations. However, a detailed discussion about the design, implementation, and benefits is beyond the scope of the current work.

There are various efforts trying to maximize SU throughputs in CRN. In [36], the coexistence problem under interference from multiple heterogeneous and independent secondary networks (SNs) was studied. An optimal coexistence strategy that can maximize the throughput achievable by an arbitrary SN was proposed. In [37], the problem of capacity maximization in CRNs was modeled as a link scheduling problem with the goal of selecting a maximum cardinality set of links including all the primary ones (with priority) and a subset of the secondary ones to transmit concurrently without causing interference to each other. All those problems were proved to be NP-hard, and polynomial time approximation algorithms were presented for better computation efficiency. However, there are evident differences between the methods of those cases and that of our scheme. For instance, both [36] and [37] focus on scheduling SUs' activities, while our scheme aims to regularize PU's activities and poses no constraints on SUs' behaviors and policies. By minimizing the PU's spectrum occupancy and the bandwidth loss due to fragmentation, we aim to maximize the amount of spare spectrum for SUs.

III. SYSTEM MODEL

We assume that the data transmissions of PUs and SUs are time slot based. For a particular spectrum band, we consider the scenario composed of a PU transmitter and multiple PU receivers. If there are more than one PU transmitters, we consider the one that is authorized for data transmission in the current time slot, either by the assignment of a negotiator or through a competition among the candidates.

As for the SUs, there is no restriction on how to utilize the idle spectrum. Thus the SUs can be organized either in a group with a centralized negotiator, or in a distributed ad-hoc manner. However, to make our discussion more tractable, we define a virtual SU on behalf of the actual SUs to perform spectrum sensing and access in the current time slot. The performance of the virtual SU, e.g., the SU throughput, is equal to the sum of the throughput of all members it represents. In this way, we can simplify the model in a time slot to be composed of one PU and one SU, and focus on the optimization of the PU traffic allocation.

As our work focuses on spectrum partition and PU traffic al-

location, the system model we adopt actually makes no assumption on either the method for spectrum sensing or the model of reporting channels. Therefore, for COP, any viable channel reporting mechanism, including the temporal dispersive reporting channels used in [38], can be employed by the SUs. However, any reporting mechanism would impose some overhead in the communications.

A. Bandwidth Partition

The entire bandwidth is denoted by B_w (in Hz) with the maximum average data rate being denoted by W (in bps) under additive white Gaussian noise (AWGN). The overall bandwidth is firstly partitioned equally into $2^M - 1$ subcarriers as $\{\alpha_0, \cdots, \alpha_{2^M-2}\}$. Then the subcarriers are non-uniformly grouped to M channels to form the set

$$B = \{\beta_0, \beta_1, \cdots, \beta_{M-1}\}, \tag{1}$$

where β_k is composed of 2^k ($k \in [0, M-1]$) subcarriers. The channel nominal capacity (in bps) is denoted by:

$$\widehat{\beta}_k = \frac{2^k}{2^M - 1} W, \ \ 0 \le k \le M-1. \tag{2}$$

Here we assume that an OFDM-like structure is employed in the physical layer. That enables the channels to be composed of both the adjacent and non-adjacent subcarriers.

The numbers of subcarriers of the channel set $\{\beta_0, \cdots \beta_{M-1}\}$ form a binary geometric sequence with the sum of $2^M - 1$. It is also possible to use other integer sequences for the spectrum partition and subcarrier grouping. For instance, Fibonacci sequence, or even more generally, any complete sequence [39] is viable for that purpose. However, it is evident that the binary geometric sequence has the least length, which is helpful to minimize the amount of sensing of SUs.

B. Effective Capacity

Under the non-fading situation, since the maximum average data rate is W (bps), we have the capacity formula as follows:

$$W = B_w \log_2(1 + \frac{\overline{P}_t}{\sigma_n^2}), \tag{3}$$

where $\overline{P}_t$ is the average input power at the transmit antenna, and σ_n^2 is the white noise power of the channel.

Now we consider the frequency selective fading effect. We assume OFDM-like structures are adopted in the physical layers of both the PU and the SU, and the fading effect is approximately flat within each subcarrier. We further assume that the average transmission power is allocated to each subcarrier. Thus, by denoting the channel gain of the ith subcarrier in the nth time slot by $h_i(n)$, the ergodic capacity of the ith subcarrier is

$$c(\alpha_i, n) = B_0 \log_2(1 + \frac{\overline{P}_t}{\sigma_n^2} |h_i(n)|^2), \tag{4}$$

where

$$B_0 = \frac{B_w}{2^M - 1} = \frac{W}{(2^M - 1)\log_2(1 + \frac{\overline{P}_t}{\sigma_n^2})}. \tag{5}$$

By assuming that the effective capacity of β_k is the sum of the ergodic capacity of all its subcarriers, we have

$$c(\beta_k, n) = \sum_{i=1}^{2^k} c(\alpha_i, n) \tag{6a}$$

$$= B_0 \sum_{i=1}^{2^k} \log_2(1 + \frac{\overline{P}_t}{\sigma_n^2} |h_i(n)|^2) \tag{6b}$$

$$= \frac{\sum_{i=1}^{2^k} \log_2(1 + \frac{\overline{P}_t}{\sigma_n^2} |h_i(n)|^2)}{2^k \log_2(1 + \frac{\overline{P}_t}{\sigma_n^2})} \widehat{\beta}_k. \tag{6c}$$

By setting

$$\eta(k, n) = 1 - \frac{\sum_{i=1}^{2^k} \log_2(1 + \frac{\overline{P}_t}{\sigma_n^2} |h_i(n)|^2)}{2^k \log_2(1 + \frac{\overline{P}_t}{\sigma_n^2})}, \tag{7}$$

we have

$$c(\beta_k, n) = (1 - \eta(k, n))\widehat{\beta}_k, \ 0 \le k \le M-1. \tag{8}$$

The value of $\eta(k, n)$ denotes the ratio of capacity loss of β_k in the nth time slot due to fading effect. Since all channels are subject to block fading, although in general $\eta(k, n)$ is time varying and channel dependent, its value can be considered to be approximately a constant for any particular channel in every time slot.

In addition, since our methods are designed to run in every time slot, in the following-up discussion, we simplify the notations $c(\beta_k, n)$ to $c(\beta_k)$, and $\eta(k, n)$ to $\eta(k)$ respectively. Thus (8) can be rewritten as:

$$c(\beta_k) = (1 - \eta(k))\widehat{\beta}_k, \ 0 \le k \le M-1. \tag{9}$$

Here we stress that the simplified expression shown above is merely for abbreviation, which will never change the time varying property of either $c(\beta_k)$ or $\eta(k)$.

As for the cases where the PU and the SU are subject to different fading effects, we further denote the capacity reduction factors of β_k for the PU and the SU by $\eta_p(k)$ and $\eta_s(k)$. Accordingly, the effective capacity of the PU and the SU are denoted by $c_p(\beta_k)$ and $c_s(\beta_k)$ respectively.

C. PU Traffic Allocation

The PU traffic allocation is channel based. A channel as a whole is either allocated to the PU entirely or remains unused. Denoting the PU traffic in any time slot under consideration by $R(n)$, and the channel subset allocated to the PU by $B_p^{(n)}$, a feasible PU traffic allocation should satisfy the following constraint:

$$c_p(B_p^{(n)}) \ge R(n). \tag{10}$$

Since our methods are designed to be executed in every time slot, the variable n is not necessarily to appear in the above equation. Thus the simplified formula can be written as follows:

$$c_p(B_p) \ge R. \tag{11}$$

Here we stress that the simplified notations do not alter the time varying property of $c(\cdot)$, R and B_p at all. The capacity loss due to bandwidth fragmentation caused by the PU traffic allocation is defined as the difference between R and its actual allocated bandwidth, which is denoted as follows.

$$\Delta(B_p) = c_p(B_p) - R. \tag{12}$$

The selection of B_p should ensure that $c_p(B_p)$ meets the bandwidth demand of R. That makes $\Delta(B_p) > 0$. In case the value of R is greater than the maximum available bandwidth, the excessive part denoted by $\Delta(R) = R - c_p(B_p)$ can be buffered, and its transmission can be postponed until the following time slot.

Here we assume that the PU transmitter has perfect feedback of CSI from the PU receiver. Thereby the PU transmitter can estimate the current effective capacity of the channel, namely $c_p(\beta_k)$, within a time slot accurately and optimizes its traffic allocation accordingly.

D. Spectrum Sensing and Access

At the beginning of each time slot, the SU senses the status of a subcarrier to determine the status of the corresponding channel. If there is no PU transmission, the SU can utilize the time slot for its own transmission. Since all the subcarriers grouped into a channel share the same occupancy status, the SU only needs to sense any selected subcarrier in each channel and infers the entire channel's occupancy status. In this way the overheads of sensing can be reduced significantly.

IV. MAXIMIZATION OF CAPACITY

We assume that the SU bandwidth demand is large, which can always use up all the spare capacity left by the PU. Therefore we have $B_s = B \setminus B_p$, where B_s is the set of all idle channels available to the SU. Our objective is to maximize $c_s(B_s)$ by optimizing B_p. As $c_s(B_s)$ represents the effective capacity of B_s, the problem is referred to as COP (*Capacity-based Optimization Problem*) and is formulated as follows:

$$\max_{B_s \subseteq B} c_s(B_s) \quad \text{s.t. } c_p(B_p) \geq R, \tag{13}$$

where $c_s(B_s) = \sum_{\beta_i \in B_s} c_s(\beta_i)$, $c_p(B_p) = \sum_{\beta_j \in B_p} c_p(\beta_j)$, and R is the PU traffic in any time slot under consideration. On the computational complexity, we have the following conclusion.

Theorem 1: COP is NP-hard.

Proof: See Appendix A. □

Since COP is an NP-hard problem, there is no effective way to find the global optimal solution to it. A viable alternative is to resort to approximation methods. Among all existing approximation algorithms, FPTAS (*Fully Polynomial-Time Approximation Scheme*) is the best type that one can hope for an NP-hard optimization problem, if assuming P $\neq$ NP [40]. To employ the FPTAS algorithm, we first introduce some definitions as follows [41].

Definition 1 (PTAS) A polynomial-time approximation scheme (PTAS) is a type of algorithm A_ε such that for each $\varepsilon > 0$, A_ε is a $(1+\varepsilon)$-approximation algorithm (for minimization problems) or a $(1-\varepsilon)$-approximation (for maximization problems).

Definition 2 (FPTAS) A fully polynomial-time approximation scheme (FPTAS) is a PTAS such that the time complexity of A_ε is bounded by a polynomial in the problem size and $1/\varepsilon$.

From $B_s = B \setminus B_p$, we have

$$c_p(B_p) = c_p(B) - c_p(B_s). \tag{14}$$

Thus (13) can be equivalently written as follows:

$$\max_{B_s \subseteq B} c_s(B_s) \quad \text{s.t. } c_p(B_s) \leq c_p(B) - R, \tag{15}$$

where $c_s(B_s) = \sum_{\beta_i \in B_s} c_s(\beta_i)$, $c_p(B_s) = \sum_{\beta_j \in B_s} c_p(\beta_j)$, and R is the PU traffic in any time slot under consideration. In the following, we use B_s^* to denote the optimal solution to (15).

For a particular channel β_i, if we define $c_s(\beta_i)$ as the value and $c_p(\beta_i)$ as the weight, (15) is obviously a 0-1 knapsack problem with the total weight limit as $c_p(B) - R$. The basic steps of the FPTAS for a knap-sack problem are as follows.

We first use a greedy algorithm with the time complexity of $O(M)$ to obtain a solution to (15). The greedy algorithm goes through all the channels of B and chooses the one with the largest $c_s(\beta_i)/c_p(\beta_i)$ iteratively, under the constraint that $c_p(B_s) \leq c_p(B) - R$. We denote the outcome by $B_s^{(g)}$. It is evident that $c_s(B_s^{(g)}) \leq c_s(B_s^*)$.

Then we set $\mu = (\varepsilon/M)c_s\big(B_s^{(g)}\big)$, where $\varepsilon > 0$ and $M = |B|$ is the total number of channels. We round each value $c_s(\beta_i)$ down to the nearest integer multiples of μ and define $c_s'(\beta_i) = \lfloor c_s(\beta_i)/\mu \rfloor$ $(0 \leq i < M)$. With the new integer value $c_s'(\beta_i)$ and the non-integer weight $c_p(\beta_i)$, a new knapsack problem as an approximation for COP is as follows.

$$\max_{B_s' \subseteq B} \{c_s'(B_s')\} \quad \text{s.t. } c_p(B_s') \leq c_p(B) - R. \tag{16}$$

The conventional method for solving a knapsack problem is based on dynamic programming. However, dynamic programming can not be employed directly, because the knapsack problem in (16) is based on a non-integer weight constraint. Here we provide an alternative dynamic programming algorithm in Algorithm 1.

On the correctness of Algorithm 1, we have the following conclusion.

Lemma 1: Algorithm 1 always gives the optimal solution to the optimization problem of (16).

Proof: See Appendix B. □

Lemma 2: The time complexity of Algorithm 1 is $O(M^2\lfloor \frac{M}{\varepsilon} \rfloor)$.

Proof: The time complexity of the first for-loop of Algorithm 1, shown in Lines 3 to 5, is $O(V)$. Since $V = \sum_{i=1}^{M} c_s'(\beta_i)$ and $c_s'(\beta_i) = \lfloor c_s(\beta_i)/\mu \rfloor$, we have $O(V) = O(M\lfloor c_s(\beta_i)/\mu \rfloor)$. For the second loop of Lines 6 to 15 the time complexity is $O(MV) = O(M^2\lfloor c_s(\beta_i)/\mu \rfloor)$. For the third one of Lines 17 to 23, because the time complexity of Line 18 is $O(V)$, the total time complexity is $O(MV) = O(M^2\lfloor c_s(\beta_i)/\mu \rfloor)$.

From all above, the time complexity of Algorithm 1 is $O(M^2\lfloor c_s(\beta_i)/\mu \rfloor)$, or equivalently $O(M^2\lfloor M/\varepsilon \rfloor)$. □

Algorithm 1 An approximation scheme for COP.

Input: M, ε, $c_s(\beta_i)$ $(0 \le i \le M-1)$
Output: $B_s'^*$

$V \leftarrow \sum_{i=1}^{M} c_s'(\beta_{i-1})$, where $\varepsilon > 0$, $\mu := \frac{\varepsilon}{M} c_s\big(B_s^{(g)}\big)$ and $c_s'(\beta_i) := \lfloor \frac{c_s(\beta_i)}{\mu} \rfloor$
$w[0,0] \leftarrow 0$
for $k := 1$ to V **do**
  $w[0,k] \leftarrow \infty$
end for
for $i := 1$ to M **do**
  for $k := 0$ to V **do**
    if $k \le c_s'(\beta_{i-1})$ **then**
      $w[i,k] \leftarrow \min\{w[i-1,k], c_p(\beta_{i-1})\}$
    else
      $w[i,k] \leftarrow \min\{w[i-1,k], c_p(\beta_{i-1})$
        $+w[i-1, k-c_s'(\beta_{i-1})]\}$
    end if
  end for
end for
$U \leftarrow c_p(B) - R$, $B_s'^* \leftarrow \emptyset$
for $i := M$ down to 1 **do**
  $k \leftarrow \arg\max_z \{w[i,z] \,|\, w[i,z] \le U\}$
  if $w[i,k] \ne w[i-1,k]$ **then**
    $U \leftarrow U - c_p(\beta_{i-1})$
    $B_s'^* \leftarrow B_s'^* \cup \beta_{i-1}$
  end if
end for
return $(B_s'^*)$

Theorem 2: Algorithm 1 is an FPTAS for COP, i.e.,

$$c_s(B_s'^*) > (1-\varepsilon) c_s(B_s^*), \tag{17}$$

where $B_s'^*$ is the output of Algorithm 1, and B_s^* is the optimal solution to COP.

The proof is omitted as it is similar to that of Theorem 3.5 in [41].

V. MAXIMIZATION OF SUBCARRIERS

Solving COP requires the PU to have the fading information of the SU. This is usually impractical, which may incur significant communication overheads and leads to implementation issues. In addition, as shown in Theorem 1, even if the SU fading information is known to the PU, COP is still an NP-hard problem. To address the complexity, we propose another surrogate problem as follows.

A. Subcarrier based Optimization Problem

If we target at maximizing the nominal capacity made available to SU, the problem reads

$$\max_{B_s \subseteq B} \widehat{B}_s \quad \text{s.t. } c_p(B_p) \ge R, \tag{18}$$

where $\widehat{B}_s = \sum_{\beta_i \in B_s} \widehat{\beta}_i$ and $c_p(B_p) = \sum_{\beta_j \in B_p} c_p(\beta_j)$. As $\widehat{\beta}_i = 2^i \widehat{\beta}_0$, the above optimization essentially maximizes the number of subcarriers for the SU, and we term it *Subcarrier based Optimization Problem* (SOP).

For SOP, we show a greedy algorithms in Algorithm 2 that achieves the global optimum with its complexity being $O(M)$, or $O(M \log M)$ if the sort operation is taken into account.

Algorithm 2 Greedy method for PU traffic allocation for SOP.

Input: M, R, $c_p(\beta_i)$ $(0 \le i \le M-1)$
Output: B_s^*

$x_i \leftarrow 1$ $(\forall i,\, 0 \le i \le M-1)$, $w \leftarrow c_p(B)$
for $i := M-1$ down to 0 **do**
  if $w - c_p(\beta_i) < R$ **then**
    continue
  else
    $w \leftarrow \big(w - c_p(\beta_i)\big)$, $x_i \leftarrow 0$
    if $w - R = 0$ **then**
      break
    end if
  end if
end for
$B_s^* \leftarrow \{\beta_i \,|\, x_i = 0,\, 0 \le i \le M-1\}$
return (B_s^*)

Theorem 3: Algorithm 2 gives the global optimal solution to SOP.

Proof: We use $N(\cdot)$ to denote the number of subcarriers of a channel or a channel set. Algorithm 2 starts its iteration from β_{M-1} and goes through the channels in descending order of the number of their subcarriers as $B = \{\beta_{M-1}, \beta_{M-2}, \cdots, \beta_0\}$, where β_{M-1} is the widest channel and β_0 the narrowest one. Because the numbers of subcarriers of all channels are all powers of two, we have, for channel set B, the following inequality for any integer $k \in [1, M-1]$:

$$N(\beta^k) > \sum_{j=0}^{k-1} N(\beta^j). \tag{19}$$

For any given R, Algorithm 2 partitions the channel set B into two disjoint subsets B_p and B_s. Suppose B_s has m elements. We sort the channels of B_s in descending order of their numbers of subcarriers: $B_s = \left\{\beta_{m-1}^{(k_{m-1})}, \beta_{m-2}^{(k_{m-2})}, \cdots, \beta_0^{(k_0)}\right\}$, where k_i $(0 \le i \le m-1)$ is the original index of $\beta_i^{(k_i)}$ in B. Thus for all i and j, $0 \le i < j \le m-1$, we have $N\big(\beta_i^{(k_i)}\big) < N\big(\beta_j^{(k_j)}\big)$.

Now we assume that there exists a better solution of l elements, denoted by

$$B_s' = \left\{\beta_{l-1}^{(k_{l-1}')}, \beta_{l-2}^{(k_{l-2}')}, \cdots, \beta_0^{(k_0')}\right\} \subseteq B,$$

which contains more subcarriers than B_s though still satisfies the PU demand R, where k_j' $(0 \le j \le l-1)$ is the original index of $\beta_j^{(k_j')}$ in B. The channels of B_s' are also sorted in descending order in the number of subcarriers.

If $k_{l-1}' \le k_{m-1}$, then by the property stated in (19), the total number of subcarriers in B_s is at least as large as that of B_s', contradicting the assumption that the latter is a better solution.

Suppose therefore $k'_{l-1} > k_{m-1}$. For solution B'_s, the capacity by channels from $B \setminus B'_s$ meets the PU demand R, otherwise B'_s is not a valid solution. Consider now the step in which Algorithm 2 examines $\beta_{l-1}^{(k'_{l-1})}$. Note that, as $k'_{l-1} > k_{m-1}$, set B_s is empty at this stage, i.e., no channel has been taken by the SU and hence all channels considered prior to k'_{l-1} have been allocated to the PU by the algorithm. Because $B \setminus B'_s \subseteq B \setminus \{\beta_{l-1}^{(k'_{l-1})}\}$, the channel subset $\{\beta_j : 0 \leq j < k'_{l-1}\}$, together with the channels allocated prior to $\beta_{l-1}^{(k'_{l-1})}$, can accommodate the PU demand R. Hence Algorithm 2 would have selected $\beta_{l-1}^{(k'_{l-1})}$ to be in B_s, contradicting that $k'_{l-1} > k_{m-1}$, and the theorem follows. □

Theorem 3 shows that SOP can be solved by a greedy algorithm, with the time complexity of $O(M \log M)$ at most.

B. Equivalence to COP under Flat Fading

Under flat fading, the PU and the SU are subject to the same percent of capacity loss for every channel. For this specific situation, we have the following conclusion.

Corollary 1: COP is equivalent to SOP under flat fading.

Proof: Under flat fading, $\eta(k)$ is uniform for all channels. From (9), the problem of (13) becomes identical to that of (18), which makes COP equivalent to SOP. □

Therefore under flat fading, the solution to SOP achieved by Algorithm 2 is exactly the global optimal solution to COP.

In practice, as the channel fading is usually frequency selective, Algorithm 2 does not necessarily bring us the optimal solution to COP. We will evaluate the actual performance of Algorithm 2 and compare it to the optimal solution to COP later in the simulation part.

VI. NUMERICAL SIMULATIONS

In this section, we evaluate the performance of the FPTAS for COP and the greedy algorithm for SOP by simulations. The effective capacity of the SU obtained by the FPTAS and the greedy algorithm is calculated and then compared with that of the optimal solution to COP. As COP is NP-hard, the exhaustive search method is adopted to find the optimal allocation of the PU traffic. The SU throughput under the conventional subcarrier equal grouping scheme with the sequential allocation approach is also presented for the purpose of comparison.

As a typical and popular subcarrier grouping scheme, equal grouping has been widely adopted in the existing literature. For example, in [42], a cyclic strategy was adopted to allocate all the subcarriers evenly to the users, with the aim of spreading each user's subcarriers as far as possible. The effect of the cyclic strategy was actually identical to dividing and grouping the subcarriers evenly. Similarly, in [19], a subcarrier selector matrix was defined for each group to divide the set of all subcarriers evenly into n groups, each of which was composed by P flat subcarriers. In [43], subcarriers were split evenly into q groups each of which was composed by n subcarriers. Compared with existing subcarrier grouping schemes, the main outcome of our scheme lies in the following aspects: 1) decreasing the sensing amount needed by SU, 2) reducing the spectrum waste due to fragmentation as much as possible. Although the equal grouping scheme can reduce the sensing amount as well, it cannot achieve 1) and 2) at the same time.

As for the implementation of the equal grouping scheme, since we pose neither restriction on its group composition, nor on the sequence of sensing, a subcarrier group may consist of both adjacent subcarriers and non-adjacent ones, and the spectrum sensing operations can be performed either sequentially or randomly. These aspects allow the equal grouping scheme to represent a broad category of spectrum allocation cases.

Some parameters for the simulations are as follows. The total licensed bandwidth in Hz is set to $B_w = 1$MHz, with its maximum data rate being $W = 7.5$Mbps. The PU traffic series is generated by following a truncated Poisson distribution on $[0, W]$ with the parameter λ. The value of λ is set to $0.35W$ and $0.65W$ for the light and heavy PU traffic patterns respectively. The number of PU traffic samples is set to be 600.

The channels are supposed to being subject to white Gaussian noise of $N(0, \sigma_n^2)$. From (3), we have the signal to noise ratio as $\overline{P_t}/\sigma_n^2 \approx 23$ dB. We use Rayleigh fading to emulate the channel fading. The probability distribution function of the channel response envelop is $f(\nu; b) = \frac{\nu}{b^2} \exp(-\nu^2/2b^2)$. Varying the value of parameter b generates different levels of fading severity. A smaller value of b makes the fading impact more severe. As $h(k, n) \sim f(\nu; b)$, we first construct the effective capacity data set for the subcarriers from (4). Then we obtain the effective capacity of the channels by following different grouping schemes.

Fig. 1 (a)–(c) show the variations of the average effective capacity of the SU with regard to the numbers of channels. The abbreviations used are as follows: 1) *COP-opt*, the exhaustive search method for COP; 2) *COP-fpt*, the FPTAS for COP; 3) *SOP-gdy*, the greedy algorithm for SOP. 4) *Eq-grp*, the conventional equal grouping together with the sequential bandwidth allocation method.

We note that all curves in Fig. 1 (a)–(c) basically follow the same trend. With the growth of M, the curves of the average effective capacity first grow rapidly, then decelerate gradually, and finally approach the maximum value. As *COP-opt* is the global optimal solution, its curve represents the upper bound for the other methods.

The reason of the zig-zag curves of *Eq-grp* lies in the capacity loss due to spectrum fragmentation caused by the equal grouping of subcarriers. As we assume that the SU always vacates the channels occupied by PU, the spectrum fragments left by the PU lead to waste of bandwidth, because such fragments can not be utilized by either the PU or the SU. With the increase of the channel number, the zig-zag behavior of the uniform scheme becomes diminishing due to the finer bandwidth partition granularity.

Under the light PU traffic ($\lambda = 0.35W$), the outcome of *COP-opt* for the largest number of channels considered is around $0.415W$, indicating the maximum value of the average effective capacity of the SU under Rayleigh fading ($b = 0.5$). The difference between the value of W and the sum of the average bandwidth consumption of the PU and the SU is due to the capacity loss caused by channel fading.

When the number of channels is small, e.g. $M \leq 6$, the value of *COP-opt* is clearly below $0.415W$. The difference is due to the capacity loss caused by bandwidth fragmentation, i.e.,

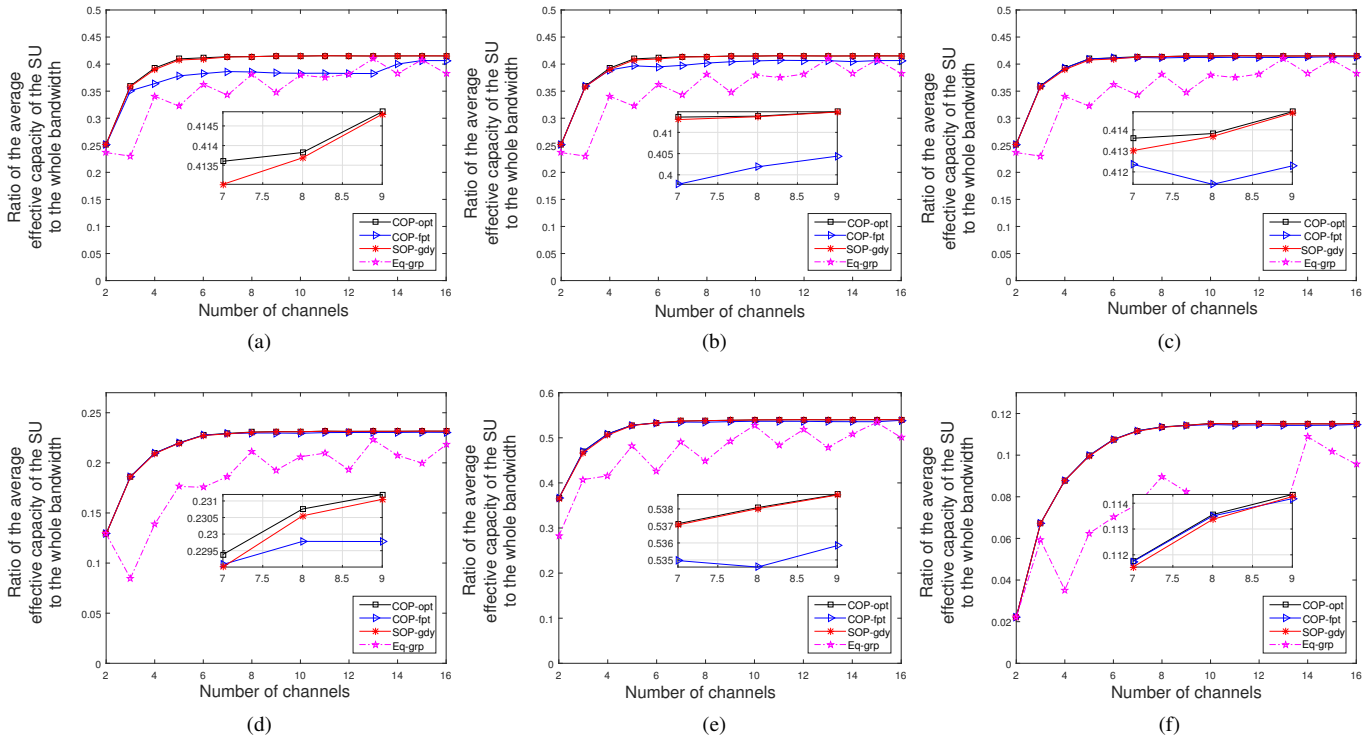


Fig. 1. Effective capacity of SU under Rayleigh fading: (a) $b = 0.5$, $\varepsilon = 0.8$, $\lambda = 0.35\mathcal{W}$, (b) $b = 0.5$, $\varepsilon = 0.4$, $\lambda = 0.35\mathcal{W}$, (c) $b = 0.5$, $\varepsilon = 0.1$, $\lambda = 0.35\mathcal{W}$, (d) $b = 0.3$, $\epsilon = 0.1$, $\lambda = 0.35\mathcal{W}$, (e) $b = 0.7$, $\epsilon = 0.1$, $\lambda = 0.35\mathcal{W}$, and (f) $b = 0.5$, $\epsilon = 0.1$, $\lambda = 0.65\mathcal{W}$.

$\Delta(B_p)$, which was defined in (12). With the growth of M, the value of $\Delta(B_p)$ decreases rapidly, and the curve of *COP-opt* rises quickly. The growing effective capacity illustrates the capacity gain obtained from the finer partition granularity. However, the capacity gain quickly becomes diminishing when M keeps growing. When $M \geq 8$, the partition granularity is quite small (i.e., $1/256W \approx 0.39\%W$) and the gain almost disappears.

Although the curve of *COP-fpt* resembles that of *COP-opt*, there are apparent gaps between the curves of *COP-fpt* and *COP-opt* in the cases of $\varepsilon = 0.8$ and $\varepsilon = 0.4$. Reducing the value of ε decreases the gap and improves the approximation accuracy accordingly. When $\varepsilon = 0.1$, the curves of *COP-fpt* and *COP-opt* are quite close.

As for *SOP-gdy*, its curve is almost overlapping with that of *COP-opt*. In the zoom-in boxes, it is evident that the gaps between *SOP-gdy* and *COP-opt* is very little and even negligible. That feature demonstrates the good performance of *SOP-gdy*. A reason for the good performance of *SOP-gdy* lies in the averaging effect. In the simulations, the effective capacity of each channel is obtained by summing up those of its subcarriers. Although the subcarrier capacity varies significantly under frequency-selective fading, the variation of the channel capacity is counterbalanced by the offsetting effect among its subcarriers. In addition, the variation is further flatted by the averaging effect over time.

Unlike the FPTAS, the greedy algorithm for SOP does not require the PU to acquire the channel fading information of the SU. Besides that, its time complexity is $O(M)$, or $O(M \log M)$ if counting into the sorting operations. This is much better than that of the FPTAS, which is at least $O(M^3)$. However, the greedy algorithm does not provide any performance guarantee.

In Fig. 1(d) and (e), we change the parameter b of Rayleigh fading from 0.5 to 0.3 and 0.7 respectively to illustrate the impact of severer or milder channel fading on the effective capacity. The parameter setting is identical to that of Fig. 1 (c) except the value of parameter b. It is clear that the general trend of the curves keeps the same, although the average effective capacity decreases or increases respectively. For the Rayleigh fading model, a larger value of the parameter b of the Rayleigh model leads to relatively less fading impact, which results in a higher SU throughput.

In Fig. 1(f), we further simulate the effective capacity under the heavy PU traffic by setting $\lambda = 0.65W$ with the rest parameters being unchanged. The curves keep the similar shape as well, except that the maximum effective capacity becomes obviously lower. All the results demonstrate the robustness of both *COP-fpt* and *SOP-gdy* with regard to various fading severity and different levels of PU traffic.

To sum up, from the simulation results, it is clear that the outputs of both the FPTAS and the greedy algorithm resemble that of the optimal solution closely. Moreover, under the non-uniform bandwidth partition scheme, the greedy algorithm is superior to the FPTAS because it does not require any fading information of the SU. In addition to that, it also has much lower time complexity than the FPTAS. In contrast, the FPTAS is superior to the greedy one for its strict approximation guarantee. By tuning the value of ε, we can get any desired approximation accuracy. Besides that, the FPTAS has a much wider applicability. It can be used not only for the particular non-uniform

bandwidth partition scheme in this paper, but also for more general multiple-channel communication cases.

However, when using the FPTAS, the number of channels does matter. A small number of channels may cause significant capacity loss due to bandwidth fragmentation, while an overly large number of channels may unnecessarily increase the time complexity. Thus a tradeoff between the accuracy and complexity needs to be considered carefully.

VII. CONCLUDING SUMMARY

In this paper, bandwidth partition and allocation are studied to optimize the spectrum utilization in cognitive communications under the interweave paradigm. We formalize the optimization of the spectrum utilization as a problem of optimizing the PU traffic allocation, namely COP. In addition, we further study the problem of maximizing the spare subcarriers, namely SOP, as the surrogate problem of COP.

By theoretical analysis and numerical simulations, the outcomes of both the FPTAS and the greedy algorithm are quite close to the optimal solution. The advantages of the FPTAS for COP include its approximation guarantee and wide applicability. The greedy algorithm is good at its high approximation performance under low time complexity and its non-requirement for fading information.

FPTAS has clearly higher complexity than the greedy algorithm; however this comes with the advantage of performance guarantee. Moreover, because the complexity is a polynomial, by algorithm theory [40], [41] it is scalable, unlike exponential-time algorithms. As for the greedy algorithm for SOP, although it performs quite well in our simulations, its effect relies heavily on the spectrum partition based on the geometric sequence, which makes its applicability highly limited.

Our discussions in this work focus on a geometric structure based non-uniform partition scheme. In our future work, the discussion would be extended to more general non-uniform partition schemes.

APPENDIX A
PROOF OF THEOREM 1

We derive a proof via a polynomial-time reduction from the partition problem [44], which is known to be NP-complete. Given a set of positive integers $a_1, a_2, \ldots, a_L$, the partition problem is to determine whether or not the set can be partitioned into two subsets with equal sum. Without any loss of generality, we can assume that the sum of all elements, $\sum_{k=1}^{L} a_k$, is an even number, because otherwise the problem has the trivial answer of no. Thus if the partition problem has a solution, then each of the two subsets has sum $\sum_{k=1}^{L} a_k/2$.

We define an instance of problem in (13) as follows. There are $M = L$ channels $0, \ldots, M-1$. For channel k, $0 \le k \le M$, $c(\beta_k) = a_k$, the PU traffic demand $R = \sum_{k=1}^{L} a_k/2$. Moreover, the special case of identical fading for PU and SU is considered, i.e., $\eta_s(k) = \eta_p(k) = \eta\,(1 \le k \le M-1)$. In this case, $c_s(\cdot)$ is simplified as $c(\cdot)$, and the expression of $c(B_s)$ reads $c(B_s) = \sum_{\beta_k \in B_s} c(\beta_k) = W - \sum_{k=0}^{M-1} \eta\widehat{\beta}_k - \sum_{\beta_k \in B_p} c(\beta_k)$. Note that the first two terms are constants (i.e., not dependent on the PU traffic allocation).

For the instance of (13) defined above, consider the following recognization version: Is there a PU traffic allocation B_p, such that the total capacity of SU is at least $W - \sum_{k=0}^{M-1} \eta\widehat{\beta}_k - \sum_{k=1}^{L} a_k/2$? Note that to satisfy the PU traffic demand, the corresponding constraint reads $\sum_{\beta_k \in B_p} a_k \ge \sum_{k=1}^{L} a_k/2$. At the same time, if $c(B_s)$ can be at least $W - \sum_{k=0}^{M-1} \eta\widehat{\beta}_k - \sum_{k=1}^{L} a_k/2$, then $\sum_{\beta_k \in B_p} a_k \le \sum_{k=1}^{L} a_k/2$. Therefore, the recognition version is equivalent to answering whether or not B_p can be chosen such that $\sum_{\beta_k \in B_p} a_k = \sum_{k=1}^{L} a_k/2$, which is in fact the partition problem. Hence the recognition version of (13) is NP-complete, and consequently the optimization version is NP-hard. □

APPENDIX B
PROOF OF LEMMA 1

Let

$$B_i = \{\beta_{j-1} \,|\, 0 < j \le i\}\ (1 \le i \le M). \tag{20}$$

B_i represents the channel set composed of the first i channels of B. We first prove that $w[i,k]$ obtained in Algorithm 1 represents the minimal weight for any channel set X ($X \subseteq B_i$) with the value being constrained by k, i.e.,

$$w[i,k] = \min_{X \subseteq B_i} \{c_p(X) \,|\, c'_s(X) \ge k\}, \tag{21}$$

where $1 \le i \le M$ and $0 \le k \le V$ ($V = \sum_{i=1}^{M} c'_s(\beta_{i-1})$). Here we prove it by induction.

For $i = 0$, we define B_0 to be the empty set. As shown in Algorithm 1, $w[0,0]$ is set to zero and $w[0,k]$ $(0 < k \le V)$ is set to ∞. Here ∞ represents the weight of a non-existing solution. Now we prove that the right hand side of (21) gives exactly the same results as that of Algorithm 1. We consider two subcases as follows.

(a) For $i = 0$ and $k = 0$, since $B_0 = \emptyset$, we have

$$\min_{X \subseteq B_0} \{c_p(X) \,|\, c'_s(X) \ge 0\} = c_p(\emptyset) = 0. \tag{22}$$

(b) For $i = 0$ and $k > 0$, since $B_0 = \emptyset$, there is no valid solution that can satisfy the non-zero value constraint. Thus we have

$$\min_{X \subseteq B_0} \{c_p(X) \,|\, c'_s(X) \ge k\} = \infty\ (\forall k \in [1, V]). \tag{23}$$

From (a) and (b), we can conclude that (21) holds for $i = 0$.

We further consider the case of $i = 1$. By following Algorithm 1, we have

$$w[1,k] = \begin{cases} \min\{w[0,0], c_p(\beta_0)\}, & k = 0; \\ \min\{w[0,k], c_p(\beta_0)\}, & 0 < k \le c'_s(\beta_0); \\ \min\{w[0,k], c_p(\beta_0) + \\ \qquad w[0, k - c'_s(\beta_0)\}, & k > c'_s(\beta_0); \end{cases} \tag{24a}$$

$$= \begin{cases} 0, & k = 0; \\ c_p(\beta_0), & 0 < k \le c'_s(\beta_0); \\ \infty, & k > c'_s(\beta_0). \end{cases} \tag{24b}$$

As for the right hand side of (21), $i = 1$ implies that $B_1 = \{\beta_0\}$. Based on that, we consider three situations as follows.

(1a) If $k = 0$, the empty set is clearly the optimal set. From $c_p(\emptyset) = 0$, we have

$$\min_{X \subseteq B_1} \{c_p(X) \,|\, c'_s(X) \geq k\}\big|_{k=0} = c_p(\emptyset) = 0. \tag{25}$$

(1b) If $0 < k \leq c'_s(\beta_0)$, there is only one set $\{\beta_0\}$ that can meet the value constraint. Thus we have

$$\min_{X \subseteq B_1} \{c_p(X) \,|\, c'_s(X) \geq k\}\big|_{0<k\leq c'_s(\beta_0)} = c_p(\beta_0). \tag{26}$$

(1c) If $k > c'_s(\beta_0)$, neither $\emptyset$ nor $\{\beta_0\}$ can satisfy the value constraint. Thus we have

$$\min_{X \subseteq B_1} \{c_p(X) \,|\, c'_s(X) \geq k\}\big|_{k>c'_s(\beta_0)} = \infty. \tag{27}$$

From (1a), (1b) and (1c), it is clear that (21) holds for $i = 1$.

Suppose (21) holds for $i = n - 1$, namely $w[n-1, k]$ is the minimal weight for any channel set X ($X \subseteq B_{n-1}$) under the value constraint of k, i.e.,

$$w[n-1, k] = \min_{X \subseteq B_{n-1}} \{c_p(X) | \, c'_s(X) \geq k\}. \tag{28}$$

Now we prove (21) holds for $i = n$, namely

$$w[n, k] = \min_{X \subseteq B_n} \{c_p(X) | \, c'_s(X) \geq k\}, \tag{29}$$

where $B_n = B_{n-1} \cup \{\beta_{n-1}\}\,(1 \leq n \leq M)$. We make the discussion under two situations as follows.

(2a) If $k \leq c'_s(\beta_{n-1})$, we further consider two subcases, i.e., whether or not β_{n-1} belongs to the optimal channel set for $i = n$. If it does, as β_{n-1} can satisfy the value constraint of k on its own and no other channels are needed, the optimal channel set must be $\{\beta_{n-1}\}$, and the minimal weight is $c_p(\beta_{n-1})$. On the contrary, if β_{n-1} is not within the optimal set, since $B_n \setminus \{\beta_{n-1}\} = B_{n-1}$, the weight optimization performed on B_n is equivalent to that on B_{n-1}, which has been defined by (28). Thus the minimal weight must be the smaller outcome of these two subcases, i.e.,

$$\min_{X \subseteq B_n} \{c_p(X) \,|\, c'_s(X) \geq k\} \tag{30a}$$
$$= \min\{\min_{X \subseteq B_{n-1}} \{c_p(X) \,|\, c'_s(X) \geq k\}, c_p(\beta_{n-1})\} \tag{30b}$$
$$= \min\{w[n-1, k], c_p(\beta_{n-1})\}. \tag{30c}$$

(2b) If $k > c'_s(\beta_{n-1})$, channel β_{n-1} can not solely match the value constraint. We again consider two subcases as follows. If β_{n-1} is within the optimal channel set for $i = n$, since β_{n-1} contributes the value of $c'_s(\beta_{n-1})$, the remaining part of the channel set has to meet the value constraint of $k - c'_s(\beta_{n-1})$. Then the optimization becomes

$$\min_{X \subseteq B_n} \{c_p(X) \,|\, c'_s(X) \geq k\} \tag{31a}$$
$$= c_p(\beta_{n-1}) + \min_{X \subseteq B_{n-1}} \{c_p(X) \,|\, c'_s(X) \geq k - c'_s(\beta_n)\} \tag{31b}$$
$$= c_p(\beta_{n-1}) + w[n-1, k - c'_s(\beta_{n-1})]. \tag{31c}$$

If β_{n-1} is not part of the optimal solution, we just make the optimization on B_{n-1} under the original value constraint, i.e.,

$$\min_{X \subseteq B_n} \{c_p(X) \,|\, c'_s(X) \geq k\} \tag{32a}$$
$$= \min_{X \subseteq B_{n-1}} \{c_p(X) \,|\, c'_s(X) \geq k\} \tag{32b}$$
$$= w[n-1, k]. \tag{32c}$$

By combining those two subcases, we have the final expression as follows.

$$\min_{X \subseteq B_n} \{c_p(X) \,|\, c'_s(X) \geq k\} = \min\{w[n-1, k], c_p(\beta_{n-1}) + w[n-1, k - c'_s(\beta_{n-1})]\}. \tag{33}$$

From the results of (2a) and (2b), it is clear that (30a) and (33) give the identical definition of $w[i, k]$ as in Algorithm 1 (Lines 9,11 and 12)). Thus (29) holds for $i = n$. By using the principle of induction, we know that (21) holds.

Next we prove that the output of Algorithm 1 is the optimal set to (16), namely $c'_s(B'^*_s)$ being the maximal value under the weight constraint of $c_p(B) - R$. We prove that by contradiction.

Suppose there exists an optimal channel set B''_s other than B'^*_s, whose value is $c'_s(B''_s)$. We assume $c'_s(B''_s)$ is better than that of B'^*_s, i.e.,

$$c'_s(B''_s) > c'_s(B'^*_s). \tag{34}$$

As all values are integers, without loss of generality, we further set

$$c'_s(B''_s) = c'_s(B'^*_s) + 1. \tag{35}$$

By following Algorithm 1 (Lines 17 to 23), we have

$$c'_s(B'^*_s) = \arg\max_z \{w[M, z] \,\big|\, w[M, z] \leq c_p(B) - R\}. \tag{36}$$

From the above equation we immediately have

$$w[M, c'_s(B'^*_s)] \leq c_p(B) - R. \tag{37}$$

By considering the monotonicity of $w[M, k]$ with respect to k, it is easy to get

$$w[M, c'_s(B'^*_s) + 1] > c_p(B) - R. \tag{38}$$

By applying (21) to $w[M, c'_s(B''_s)]$, we have

$$w[M, c'_s(B''_s)] = \min_{X \subseteq B} \{c_p(X) \,|\, c'_s(X) \geq c'_s(B''_s)\}. \tag{39}$$

From the above equation, we immediately have

$$c_p(B''_s) \geq w[M, c'_s(B''_s)] \tag{40a}$$
$$= w[M, c'_s(B'^*_s) + 1] \tag{40b}$$
$$> c_p(B) - R. \tag{40c}$$

It is clear that B''_s violates the weight constraint. Thus the assumption of (34) must be incorrect, and the result follows. □

REFERENCES

[1] Federal Communications Commission: FCC, ET Docket No. 03-289 "Notice of inquiry and notice of proposed Rulemaking," Tech. Rep., Nov. 2003.

[2] J. Mitola and G. Q. Maguire Jr., "Cognitive radio: Making software radios more personal," *IEEE Pers. Commun*, vol. 6, no. 4, pp. 13–18, Aug. 1999.

[3] 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, Sept. 2006.

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

[5] Y.-C. Liang, K.-C. Chen, G. Y. Li, and P. Mahonen, "Cognitive radio networking and communications: An overview," *IEEE Trans. Veh. Technol.*, vol. 60, no. 7, pp. 3386–3407, Sept. 2011.

[6] A. S. Alfa, B. T. Maharaj, S. Lall, and S. Pal, "Mixed-integer programming based techniques for resource allocation in underlay cognitive radio networks: A survey," *J. Commun. Netw.*, vol. 18, no. 5, pp. 744–761, Oct. 2016.

[7] Y. Wang, Y. Xu, L. Shen, C. Xu, and Y. Cheng, "Two-dimensional POMDP-based opportunistic spectrum access in time-varying environment with fading channels," *J. Commun. Netw.*, vol. 16, no. 2, pp. 217–226, Apr. 2014.

[8] S. Wang, B. T. Maharaj, and A. S. Alfa, "A maximum throughput channel allocation protocol in multi-channel multi-user cognitive radio network," *J. Commun. Netw.*, vol. 20, no. 2, pp. 111-121, Apr. 2018.

[9] H. Kim, G. P. Villardi, and J. Ma, "Energy efficient dynamic resource allocation for wireless systems using receiver puncturing technique," *J. Commun. Netw.*, vol. 20, no. 4, pp. 366–373, Aug. 2018.

[10] S. Huang, A. Ephremides, and D. Yuan, "Improving the cognitive access efficiency by non-uniform bandwidth allocation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6435–6447, Nov. 2015.

[11] M. D. P. Guirao, A. Wilzeck, A. Schmidt, K. Septinus, and C. Thein, "Locally and temporary shared spectrum as opportunity for vertical sectors in 5g," *IEEE Netw.*, vol. 31, no. 6, pp. 24–31, Nov. 2017.

[12] W. Dong, S. Rallapalli, L. Qiu, K. Ramakrishnan, and Y. Zhang, "Double auctions for dynamic spectrum allocation," *IEEE/ACM Trans. Netw.*, vol. 24, no. 4, pp. 2485–2497, Oct. 2016.

[13] M. Tao and Y. Liu, "Spectrum leasing and cooperative resource allocation in cognitive OFDMA networks," *J. Commun. Netw.*, vol. 15, no. 1, pp. 102–110, Feb. 2013.

[14] R. I. C. Chiang, G. B. Rowe, and K. W. Sowerby, "A quantitative analysis of spectral occupancy measurements for cognitive radio," in *Proc. IEEE VTC*, Apr. 2007, pp. 3016–3020.

[15] F. K. Jondral, "Software-defined radio: Basics and evolution to cognitive radio," *EURASIP J. Wireless Commun. Netw.*, vol. 2005, no. 3, pp. 275–283, Aug. 2005.

[16] I. F. Akyildiz, W. Y. Lee, M. C. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Commun. Mag.*, vol. 46, no. 4, pp. 40–48, Apr. 2008.

[17] E. Coffman, P. Robert, F. Simatos, S. Tarumi, and G. Zussman, "Channel fragmentation in dynamic spectrum access systems: A theoretical study," in *Proc. ACM SIGMETRICS*, June 2010, pp. 333–344.

[18] 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.

[19] Z. Liu, Y. Xin, and G. B. Giannakis, "Linear constellation precoding for ofdm with maximum multipath diversity and coding gains," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 416–427, Apr. 2003.

[20] F. Jin, G. Sahin, A. Arora, and H. A. Choi, "The effects of the subcarrier grouping on multi-carrier channel aware scheduling," in *Proc. BroadNets*, Oct. 2004, pp. 632–640.

[21] N. Khambekar, L. Dong, and V. Chaudhary, "Utilizing ofdm guard interval for spectrum sensing," in *Proc. IEEE Wireless Commun. Netw.*, Mar. 2007, pp. 38–42.

[22] W. Hu, D. Willkomm, M. Abusubaih, J. Gross, G. Vlantis, M. Gerla, and A. Wolisz, "Cognitive radios for dynamic spectrum access-dynamic frequency hopping communities for efficient ieee 802.22 operation," *IEEE Commun. Mag.*, vol. 45, no. 5, pp. 80–87, May 2007.

[23] N. S. Shankar, C. Cordeiro, and K. Challapali, "Spectrum agile radios: Utilization and sensing architectures," in *Proc. IEEE DySPAN*, Nov. 2005, pp. 160–169.

[24] A. Ghasemi and E. S. Sousa, "Optimization of spectrum sensing for opportunistic spectrum access in cognitive radio networks," in *Proc. IEEE Consum. Commun. Netw.*, Jan. 2007, pp. 1022–1026.

[25] D. Datla, R. Rajbanshi, A. M. Wyglinski, and G. J. Minden, "Parametric adaptive spectrum sensing framework for dynamic spectrum access networks," in *Proc. IEEE DySPAN*, Apr. 2007, pp. 482–485.

[26] S. Huang, A. Ephremides, and D. Yuan, "Optimal allocation of non-uniformly partitioned bandwidth for cognitive communications under fading conditions," in *Proc. IEEE ICC*, May 2016, pp. 1–6.

[27] K. Tan, J. Fang, Y. Zhang, S. Chen, L. Shi,J. Zhang, and Y. Zhang, "Fine-grained channel access in wireless LAN," in *Proc. ACM SIGCOMM*, Aug. 2010, pp. 147–158.

[28] S. Rayanchu, V. Shrivastava, S. Banerjee, R. Chandra, "FLUID: Improving throughputs in enterprise wireless lans through flexible channelization," *IEEE Trans. Mobile Comput.*, vol. 11, no. 9, pp. 1455–1469, Sept. 2012.

[29] S. S. Hong, J. Mehlman, and S. Katti, "Picasso: Flexible RF and spectrum slicing," in *Proc. ACM SIGCOMM* , 2012, pp. 37–48.

[30] L. Jiao, F .Y. Li, and V. Pla, "Modeling and performance analysis of channel assembling in multichannel cognitive radio networks with spectrum adaptation," *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2686–2697, July 2012.

[31] L. Jiao, I. A. Balapuwaduge, F. Y. Li, and V. Pla, "On the performance of channel assembling and fragmentation in cognitive radio networks, " *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5661–5675, Oct. 2014.

[32] Y. Chen and H-S. Oh, "A Survey of Measurement-Based Spectrum Occupancy Modeling for Cognitive Radios," *IEEE Commun. Surv. Tuts.*, vol. 18, no. 1, pp. 848–859, Jan. 2016.

[33] X. Xing, T. Jing, W. Cheng, Y. Huo, and X. Cheng, "Spectrum prediction in cognitive radio networks," *IEEE Wireless Commun.*, vol. 20, no. 2, pp. 90–96, Apr. 2013.

[34] Y. Saleem and M. H. Rehmani, "Primary radio user activity models for cognitive radio networks: A survey," *J. Netw. Comput. Appl.*, vol. 43, pp. 1–16, Aug. 2014.

[35] M. Höyhtyä, A. Mämmelä, M. Eskola, et al., "Spectrum Occupancy Measurements: A Survey and Use of Interference Maps," *IEEE Commun. Surv. Tuts.*, vol. 18, no. 4, pp. 2386–2414, Nov. 2016.

[36] A. S. Cacciapuoti, M. Caleffi, and L. Paura, "Optimal strategy design for enabling the coexistence of heterogeneous networks in TV white space," *IEEE Trans. Veh. Technol.*, vol. 65, no. 9, pp. 7361–7373, Sept. 2016.

[37] M. Brown, C. Marshall, D. Yang, M. Li, J. Lin, and G. Xue, "Maximizing capacity in cognitive radio networks under physical interference model," *IEEE/ACM Trans. Netw.*, vol. 25, no. 5, pp. 3003–3015, Oct. 2017.

[38] A. S. Cacciapuoti, M. Caleffi, D. Izzo, and L. Paura, "Cooperative spectrum sensing techniques with temporal dispersive reporting channels, " *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3392–3402, Oct. 2011.

[39] J. L. Brown, "Note on complete sequences of integers, " *Amer. Math. Monthly.*, vol. 68, no. 6, pp. 557–560, 1961.

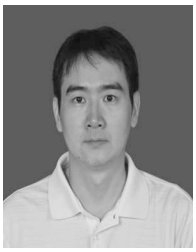
[40] V. V. Vazirani, *Approximation algorithms*, Springer Science & Business Media, 2013.

[41] D. P. Williamson and D. B. Shmoys, *The design of approximation algorithms*, Cambridge, U.K.: Cambridge Univ. Press, 2011.

[42] Z. Wang and G. B. Georgios, "Wireless multicarrier communications," *IEEE Signal Process Mag.*, vol. 17, no. 3, pp. 29–48, 2000.

[43] N. Prasad, L. Venturino, and X. Wang, "Diversity-multiplexing tradeoff analysis for OFDM systems with subcarrier grouping, linear precoding, and linear detection," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 6078–6096, 2010.

[44] M. Garey and D. Johnson, "Computers and intractability: A guide to the theory of NP-completeness", in *W. H. Freeman*, San Fr, 1979.



Song Huang received his PhD degree in Communications and Information Systems from South China University of Technology (SCUT) in 2007. He is a lecturer with the School of Computer Science and Engineering, SCUT. From Nov. 2013 to Dec. 2014, he was a visiting researcher at University of Maryland, College Park, MD, USA. His research interests include cognitive communication and network information theory.

Di Yuan received his MSc degree in Computer Science and Engineering, and PhD degree in Optimization at Linköping Institute of Technology in 1996 and 2001, respectively. He is full professor in telecommunications at the Department of Science and Technology, Linköping University, and head of a research group in mobile telecommunications. At present he is Visiting Professor at University of Maryland, College Park, MD, USA. His current research mainly addresses network optimization of 4G and 5G systems, and capacity optimization of wireless networks. Dr Yuan has been guest professor at the Technical University of Milan (Politecnico di Milano), Italy, in 2008, and senior visiting scientist at Ranplan Wireless Network Design Ltd, United Kingdom, in 2009 and 2012. In 2011 and 2013 he has been part time with Ericsson Research, Sweden. He is an area editor of the Computer Networks journal. He has been in the management committee of four European Cooperation in field of Scientific and Technical Research (COST) actions, invited lecturer of European Network of Excellence EuroNF, and Principal Investigator of several European FP7 and Horizon 2020 projects. He is a co-recipient of IEEE ICC'12 Best Paper Award, and supervisor of the Best Student Journal Paper Award by the IEEE Sweden Joint VT-COM-IT Chapter in 2014. He is a Senior Member of IEEE.

Anthony Ephremides holds the Cynthia Kim Professorship of Information Technology at the Electrical and Computer Engineering Department of the University of Maryland in College Park where he is a Distinguished University Professor and has a joint appointment at the Institute for Systems Research, of which he was among the founding members in 1986. He obtained his PhD in Electrical Engineering from Princeton University in 1971 and has been with the University of Maryland ever since. He has been recently named Distinguished University Professor.

He has held various visiting positions at other Institutions (including MIT, UC Berkeley, ETH Zurich, INRIA, etc) and co-founded and co-directed a NASA-funded Center on Satellite and Hybrid Communication Networks in 1991. He has been the President of Pontos, Inc, since 1980 and has served as President of the IEEE Information Theory Society in 1987 and as a member of the IEEE Board of Directors in 1989 and 1990. He has been the General Chair and/or the Technical Program Chair of several technical conferences (including the IEEE Information Theory Symposium in1991, 2000, and 2011,the IEEE Conference on Decision and Control in 1986, the ACM Mobihoc in 2003, and the IEEE Infocom in 1999). He has served on the Editorial Board of numerous journals and was the Founding Director of the Fairchild Scholars and Doctoral Fellows Program, a University-Industry Partnership from 1981 to 1985.

He has received the IEEE Donald E. Fink Prize Paper Award in 1991 and the first ACM Achievement Award for Contributions to Wireless Networking in 1996, as well as the 2000 Fred W. Ellersick MILCOM Best Paper Award, the IEEE Third Millennium Medal, the 2000 Outstanding Systems Engineering Faculty Award from the Institute for Systems Research, and the Kirwan Faculty Research and Scholarship Prize from the University of Maryland in 2001, and a few other official recognitions of his work. He also received the 2006 Aaron Wyner Award for Exceptional Service and Leadership to the IEEE Information Theory Society.

He is the author of several hundred papers, conference presentations, and patents, and his research interests lie in the areas of Communication Systems and Networks and all related disciplines, such as Information Theory, Control and Optimization, Satellite Systems, Queueing Models, Signal Processing, etc. He is especially interested in Wireless Networks and Energy Efficient Systems.