# A Decomposition Approach to Quality-Driven Multiuser Video Streaming in Cellular Cognitive Radio Networks

Zhifeng He, *Student Member, IEEE*, Shiwen Mao, *Senior Member, IEEE*, and Sastry Kompella, *Senior Member, IEEE*

***Abstract*—We tackle the challenging problem of streaming multiuser videos over the downlink of a cellular cognitive radio network (CRN), where each cognitive user (CU) can sense and access multiple channels at a time. Spectrum sensing, channel assignment, and power allocation strategies are jointly optimized to maximize the quality of service (QoS) for the CUs. We show that the formulated mixed integer nonlinear programming (MINLP) problem can be decomposed into two subproblems: 1) SP1 for the optimal spectrum sensing strategy and 2) SP2 for the optimal channel assignment and power allocation, without sacrificing optimality. We show that SP1 can be optimally solved if there is no restriction on the sensing capability for each CU, and develop a column generation (CG)-based algorithm to solve SP2 iteratively in a distributed manner. We also develop a heuristic algorithm for spectrum sensing with greatly reduced requirement on CU hardware, while still achieving a highly competitive sensing performance. We analyze the proposed algorithms with respect to complexity and time efficiency, and derive a performance upper bound. The proposed algorithms are validated with simulations.**

***Index Terms*—Quality of service, cognitive radio networks, column generation, multiuser video streaming, optimization.**

## I. Introduction

A RECENT study by Cisco reveals an unprecedented paradigm change in wireless data: there is not only a dramatic growth in volume, but also a fundamental change in the composition of wireless traffic, a majority of which will be video related in the near future [1]. Such fast growth in wireless video traffic, coupled with the depleting spectrum resource, calls for a more flexible management of radio resources in today's and future wireless networks, in order to unlock the wireless network capacity by promoting more efficient use of spectrum.

Manuscript received December 3, 2014; revised May 21, 2015; accepted September 2, 2015. Date of publication September 10, 2015; date of current version January 7, 2016. This work was supported in part by the U.S. National Science Foundation (NSF) under Grant CNS-0953513, and in part by the Wireless Engineering Research and Education Center (WEREC), Auburn University. This work was presented in part at the 11th IEEE International Conference on Mobile Ad Hoc and Sensor Systems, Philadelphia, PA, USA, October 2014. The associate editor coordinating the review of this paper and approving it for publication was Dr. Yan Xin.

Z. He and S. Mao are with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201 USA (e-mail: zzh0008@tigermail.auburn.edu; smao@ieee.org).

S. Kompella is with the Information Technology Division, U.S. Naval Research Laboratory, Washington, DC 20375 USA (e-mail: sk@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2015.2477509

To this end, the Cognitive Radio (CR) technology has been widely recognized as an effective solution for efficient and flexible access to the radio spectrum [2]. In a CR network (CRN), cognitive users (CU) detect primary user (PU) activities on licensed channels, and access the channels that are sensed idle opportunistically and unobtrusively [3]. Bandwidth-demanding and elastic mobile services, such as wireless video, will benefit enormously from this new wireless networking paradigm [4].

Although with great potential, the problem of video over CRNs brings about a whole level of technical challenges, particularly due to the extra dimension of dynamics on channel availability and the uncertainty from spectrum sensing and access. The manifold design trade-offs, multifarious network dynamics, limited network resources and, on the other hand, video's stringent QoS constraints, necessitate a holistic cross-layer design approach to "squeeze" the most out of the CRN. Usually such cross-layer design results in a tremendously complex global optimization problem, where all the layers (i.e., the PHY, MAC, network, and application layers) and all the users (i.e., PUs and CUs) are tightly coupled [5], [6]. A *decomposition principle* that helps to decouple the design of spectrum sensing, access, and application QoS provisioning would be crucial for making the problem manageable [10].

In this paper, we tackle the problem of downlink multi-user video streaming in a cellular CRN, where each CU receives a video stream from the Cognitive Base Station (CBS). We consider the general case that each CU is able to sense (with multiple sensors or sequentially sensing the channels [11]) and access (e.g., with channel bonding/aggregation [12]) multiple channels. We adopt the H.264 Scalable Video Coding (SVC) (Quality Scalability) model from [14], and jointly design spectrum sensing, channel access, and power control for maximizing the QoS of all the CUs. There are two tightly coupled parts in this problem: the spectrum sensing problem (SP1) to determine which CU to sense which channel; and the channel assignment and power allocation problem (SP2) to allocate channels and transmit power to the CUs.

The formulated problem turns out to be a Mixed Integer NonLinear Programming (MINLP) problem, which is NP-hard in general. However, as in [10], where a *separation principle* is established to decouple the design of sensing strategy from that of sensor and access policy, we show that our problem can also be decoupled into two relatively easier sub-problems with a *decomposition principle* and develop an effective Column Generation (CG) based solution algorithm [15].

The major contributions made in this paper include:

1) A holistic *problem formulation* that jointly optimizes the spectrum sensing, channel assignment, and power allocation strategies for maximizing CU QoS.
2) A *decomposition principle* to decouple the original problem into a sensing strategy optimization problem SP1 and a resource allocation problem SP2, without sacrificing optimality if there is no restriction on the sensing capability for each CU, and *effective algorithms* to solve SP1 and SP2.
3) A *heuristic sensing scheme* that is less demanding on CU hardware than the optimal sensing strategy, but can achieve highly competitive sensing performance.
4) An *upper bound* for the performance of the CG-based distributed algorithm and an analysis of complexity and efficiency in terms of time savings.
5) *Simulation validation* to demonstrate the superior performance of our proposed algorithms in terms of sensing performance and the QoS achieved by CUs.

The remainder of this paper is organized as follows. Section II reviews related work. The system model and problem formulation are presented in Section III, while the decomposition principle and the two sub-problems SP1 and SP2 are presented in Section IV. The CG-based distributed algorithm to solve SP2 is developed in Section V and analyzed in Section VI. Section VII presents the performance validation and Section VIII concludes the paper.

## II. Related Work

CR research has been largely focused on the aspects of spectrum sensing and dynamic spectrum access. In [16], the authors study the sensing-throughput tradeoff problem that optimizes the spectrum sensing time so that the CU's throughput can be maximized with restricted interference to the PUs. Unlike [16], the protocol proposed in [10] also considers the problem of which channel to sense, in addition to sensing parameters and access strategy optimization. Moreover, the design of sensing strategy is independent to sensing parameters design and the access strategy, as specified in a *principle of separation* [10]. These works focus on the optimization of sensing parameters only, and there is no collaboration among CUs. Considering the fact that different CUs may have different spectrum sensing performance, the authors in [17] propose an algorithm where groups of CUs are formed for cooperative sensing, aiming to find the best grouping scheme to discover most idle channels. Furthermore, the problem of sensing parameter optimization in addition to optimal sensor selection is addressed in [18], in order to achieve a trade-off between detection performance and sensing overhead.

Recently, cross layer design for video streaming over CRNs has attracted considerable interest. Unlike exploiting new spectrum frontiers [19], the aim is to make more efficient use of allocated spectrum for enhanced video service. An auction game model is proposed in [20] to solve the problem of spectrum allocation in delay-sensitive content-aware multimedia delivering. Channel/path selection for multi-user video streaming is formulated as an MINLP problem in [5] to maximize the received video quality while restricting collisions with PUs. Packet scheduling is studied in [12] in which spectrum sensing at the PHY is integrated with packet scheduling at the MAC layer to improve delay-QoS provisioning over CRNs. The authors also analyze the throughput and delay performance with a Markov chain and $M/G_Y/1$ queuing model. Beyond these, other cross layer factors such as Fine Grained Scalability (FGS) coding, error control, and modulation, are jointly considered in [4] to achieve the maximum QoS for CUs in a cellular CRN. Interestingly, cross layer optimization of streaming videos over a CR link can also be modeled as a POMDP (Partially Observable Markov Decision Process) as in [21], in which intra refreshing rate, a video codec parameter, along with spectrum sensing and access strategies are jointly designed.

Different from the above mentioned prior works that consider the physical layer factors, some other works treat the problem of video streaming over CRNs in the upper layers. The authors in [7] model the routing problem in video streaming over CRNs as a decision tree problem. Then the routing scheme is optimized in order to maximize the PSNR of the received video sequence. Considering the fact that different CUs may have different channel data rate and different buffer storage size, which result in different abilities of tolerating network dynamics, the authors of [8] propose to allocate channels to CUs according to their buffer storage. Reducing the playout speed to a certain extent when the buffered data at the receiver is low, is also a feasible solution to reduce the probability of playout outage, as the authors propose in [9].

This paper is motivated by these interesting prior works, and is mainly focused on the joint design of spectrum sensing and resource allocation strategies for streaming multi-user videos over the downlink of a cellular CRN with a novel decomposition principle, which is not well addressed in prior work but is essential for supporting the demand of large bandwidth for video applications and enhance the QoS of CUs.

## III. System Model and Problem Statement

### A. System Model

We consider a primary network operating on $N_1$ licensed orthogonal channels, while each channel $j$ has bandwidth $B_j$. A CR network is co-located with the primary network, consisting of a CBS and $M$ CUs. The CUs sense the PU activities on the licensed channels and access the channels in an opportunistic manner. As in [11], we first assume that each CU is equipped with $N_1$ sensors so that it can sense all the channels simultaneously. This assumption is relaxed in Section IV-B, where each CU can only sense a few channels at a time. The CBS determines the status of the licensed channels based on the sensing results reported from the CUs.

We consider the scenario of downlink multi-user video streaming, where the CBS transmits different video streams to the CUs using the channels sensed idle. Once the channel states are estimated, the CBS and CUs determine the allocation of the idle channels, and the CBS selects a power level $k$,

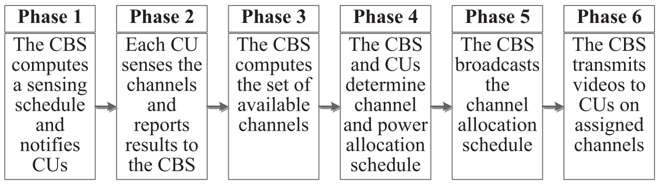


Fig. 1. Operations of CBS and CUs in a time slot.

$k = 1, 2, \ldots, K$, for the video transmission to a CU on each allocated channel. We assume that each CU and the CBS adopt the channel bonding/aggregation technique [12], [13], such that they can transmit on multiple assigned channels simultaneously to make use of all the available spectrum. To enforce a certain level of fairness among the CRs, we define an upper bound $C_i$ on how much total time a CU $i$ can access all the channels. If $C_i$ is less than the total number of channels, this can limit how much channel time a CR can have at most. Otherwise, if $C_i$ is equal to the number of channels, then there is no such fairness constraint and a CR can access all the channels for the entire time slot. We assume time is divided into a series of non-overlapping Group of Pictures (GOP) windows, each consisting of $\mathbb{T}$ time slots. The operations of the CBS and CUs in each time slot, as discussed above, are summarized in Fig. 1.

### B. Problem Statement

Due to multipath fading and shadowing, different CUs usually experience different SNRs when detecting a PU signal, and thus may have different sensing performance. It is important to choose a suitable set of CUs to sense a licensed channel [17]. To achieve diversity gain, cooperative sensing is usually used to improve the detection performance by fusing the sensing results from multiple CUs, where a certain fusion rule is used to combine the CU sensing results. In this paper, we consider cooperative sensing with the OR fusion rule: if any of the CUs reports the presence of a PU signal on a channel, the CBS will determine that the channel is busy; otherwise, the channel is considered to be idle.

We use an $M \times N_1$ matrix $\mathbf{X}$ to represent the assignment of sensing tasks, where each element $x_{ij}$ is defined as

$$x_{ij} = \begin{cases} 1, & \text{CU } i \text{ is assigned to sense channel } j \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

For energy detection, the probability of detection of PU signal on channel $j$ by CU $i$, $P_{d_{ij}}$, and the probability of false alarm on channel $j$ by CU $i$, $P_{f_{ij}}$, can be expressed as [16]

$$\begin{cases} P_{d_{ij}} = \frac{1}{2}\text{erfc}\left(\left(\frac{\lambda_{ij}}{n_0 B_j} - \gamma_{ij} - 1\right)\sqrt{\frac{\kappa}{2(2\gamma_{ij}+1)}}\right) \\ P_{f_{ij}} = \frac{1}{2}\text{erfc}\left(\sqrt{2\gamma_{ij}+1}\ \text{erfc}^{-1}\left(2P_{d_{ij}}\right) + \sqrt{\frac{\kappa}{2}}\gamma_{ij}\right), \end{cases} \quad (2)$$

where $n_0$ is the power density of Additive White Gaussian Noise (AWGN) which is i.i.d. at each CU, $\gamma_{ij}$ is the SNR of PU signal on channel $j$ at CU $i$'s side, $\lambda_{ij}$ is the energy detection threshold on channel $j$ by CU $i$, and $\kappa$ is the number of samples collected. For cooperative sensing with the OR fusion rule, the probability of detection of PU signal on channel $j$, $P_{d_j}$, and the probability of false alarm on channel $j$, $P_{f_j}$, can be written as

$$P_{d_j} = 1 - \prod_{i=1}^{M}(1 - P_{d_{ij}})^{x_{ij}}, \ \ P_{f_j} = 1 - \prod_{i=1}^{M}(1 - P_{f_{ij}})^{x_{ij}}. \quad (3)$$

To provide a graceful protection to PUs, we set $P_{d_j} = P_{req}$, where $P_{req}$ is the maximum interference from the CU system that can be tolerated by the PU system, and $\sum_{i=1}^{M} x_{ij} = \Lambda_j$, for all $j = 1, 2, \ldots, N_1$, where $\Lambda_j$ is the minimum number of CUs to sense channel $j$. If we set $P_{d_{ij}} = \bar{P}_{d_j} = 1 - (1 - P_{req})^{1/\Lambda_j}$, it follows (3) that $P_{d_j} = P_{req}$.

Let $r_j$ denote the real state of channel $j$: $r_j = 0$ when channel $j$ is idle, and $r_j = 1$ otherwise. Also let $s_j$ be the cooperative sensing result on channel $j$: $s_j = 0$ if the channel is determined idle, and $s_j = 1$ otherwise. We have

$$\begin{cases} P(s_j = 0) = P(r_j = 0)(1 - P_{f_j}) + P(r_j = 1)(1 - P_{d_j}) \\ P(s_j = 1) = P(r_j = 0)P_{f_j} + P(r_j = 1)P_{d_j}. \end{cases} \quad (4)$$

The cooperative sensing results on the $N_1$ channels can be represented as $\vec{\mathbf{S}} = \left\{s_j, j = 1, \ldots, N_1\right\}$. There are $2^{N_1}$ possible outcomes for $\vec{\mathbf{S}}$, and let $\vec{\mathbf{S}}_h$ be the $h$-th outcome, $0 \leq h \leq 2^{N_1} - 1$. To determine the $j$-th element in $\vec{\mathbf{S}}_h$, let $s_j = \Gamma_j(h)$, $j = 1, 2, \ldots, N_1$, denote the relationship between $\vec{\mathbf{S}}_h$ and $s_j$. Assuming independent channel states, the probability of getting outcome $\vec{\mathbf{S}}_h$ can be written as

$$\begin{aligned} P(\vec{\mathbf{S}} = \vec{\mathbf{S}}_h) &= \prod_{j=1}^{N_1} P(s_j = \Gamma_j(h)) \\ &= \prod_{j=1}^{N_1} \left[(1 - \Gamma_j(h))P(s_j = 0) + \Gamma_j(h)P(s_j = 1)\right]. \end{aligned} \quad (5)$$

We adopt the QoS model for H.264 SVC (Quality Scalability) from [14] as

$$\rho_i = \alpha_i + \beta_i \cdot R_i, \quad (6)$$

where $\rho_i$ is the Y-PSNR (Peak Signal-to-Noise Ratio) of the received video at CU $i$, $\alpha_i$ and $\beta_i$ are constants dependent on the content type of the video sequence, and $R_i$ is the effective data rate of the video sequence. According to conditional expectation, the expected overall QoS can be derived as

$$\begin{aligned} \mathrm{E}\left(\sum_{i=1}^{M}\rho_i\right) &= \sum_{i=1}^{M}\sum_{h=0}^{2^{N_1}-1} \mathrm{E}(\rho_i | \vec{\mathbf{S}} = \vec{\mathbf{S}}_h) P(\vec{\mathbf{S}} = \vec{\mathbf{S}}_h) \\ &= \sum_{h=0}^{2^{N_1}-1}\sum_{i=1}^{M} \mathrm{E}(\rho_i | \vec{\mathbf{S}} = \vec{\mathbf{S}}_h) P(\vec{\mathbf{S}} = \vec{\mathbf{S}}_h). \end{aligned} \quad (7)$$

For $\vec{\mathbf{S}} = \vec{\mathbf{S}}_h$, let $\Phi_h = \left\{j : \Gamma_j(h) = 0, j = 1, 2, \ldots, N_1\right\}$ be the set of channels sensed idle. Let $\mathbf{Y}_h = \left[y_{ijk}^h\right]$, $1 \leq i \leq M$, $j \in \Phi_h$, $1 \leq k \leq K$, be the channel assignment and power allocation matrix, where $0 \leq y_{ijk}^h \leq 1$ is the amount of time that

CBS transmits to CU $i$ with a power level of $k$ on channel $j$ in a time slot, when the sensing outcome is $\vec{\mathbf{S}}_h$. The channel assignment and power allocation strategy can be expressed as $\mathbf{Y} = \left[\mathbf{Y}_0, \mathbf{Y}_1, \ldots \mathbf{Y}_{2^{N_1}-1}\right]$.

Putting it all together, it follows that

$$\begin{aligned} \mathrm{E}(\rho_i|\vec{\mathbf{S}} = \vec{\mathbf{S}}_h) &= \alpha_i + \beta_i \cdot \mathrm{E}(R_i|\vec{\mathbf{S}} = \vec{\mathbf{S}}_h) \\ &= \alpha_i + \beta_i \cdot \mathrm{E}\left(\sum_{j\in\Phi_h}\sum_{k=1}^{K} R_{ijk} \cdot y_{ijk}^h|\vec{\mathbf{S}} = \vec{\mathbf{S}}_h\right) \\ &= \alpha_i + \beta_i \sum_{j\in\Phi_h}\sum_{k=1}^{K}\left(P_{00}^j R_{00}^{ijk} + P_{10}^j R_{10}^{ijk}\right) \cdot y_{ijk}^h \end{aligned} \tag{8}$$

where $G_k$ is the power of level $k$, $d_{ij}$ is the channel gain between the CBS and CU $i$ on channel $j$, $P_{00}^j = P(r_j = 0|s_j = 0) = \frac{(1-P_{f_j})P(r_j=0)}{(1-P_{f_j})P(r_j=0)+(1-P_{d_j})P(r_j=1)}$, $P_{10}^j = P(r_j = 1|s_j = 0) = 1 - P_{00}^j$, $R_{00}^{ijk} = B_j \log_2(1 + G_k d_{ij}/(n_0 B_j))$, and $R_{10}^{ijk} = B_j \log_2(1 + G_k d_{ij}/(n_0 B_j(1 + \gamma_{ij})))$.

Define $w_{ijk} = \alpha_i + \beta_i B_j(P_{00}^j R_{00}^{ijk} + P_{10}^j R_{10}^{ijk})$. The master problem of maximizing the total expected QoS, denoted as P0, can be formulated as follows.

$$\underline{\mathbf{P0}} : \max : \sum_{h=0}^{2^{N_1}-1}\sum_{i=1}^{M}\sum_{j\in\Phi_h}\sum_{k=1}^{K} w_{ijk} \cdot y_{ijk}^h \cdot P(\vec{\mathbf{S}} = \vec{\mathbf{S}}_h) \tag{9}$$

$$\text{s.t.} \quad \sum_{j\in\Phi_h}\sum_{k=1}^{K} y_{ijk}^h \le C_i, \forall\, i, h \tag{10}$$

$$\sum_{i=1}^{M}\sum_{k=1}^{K} y_{ijk}^h \le 1, \forall\, j \tag{11}$$

$$\sum_{i=1}^{M}\sum_{j\in\Phi_h}\sum_{k=1}^{K} y_{ijk}^h \cdot G_k \le G_{total}, \forall\, h \tag{12}$$

$$\sum_{i=1}^{M} x_{ij} = \Lambda_j, \forall\, j \tag{13}$$

$$x_{ij} = \{0, 1\}, \forall\, i, j \tag{14}$$

$$y_{ijk}^h \begin{cases} \in [0, 1], & \text{if } G_k d_{ij}/(n_0 B_j) \ge \bar{\gamma} \\ = 0, & \text{otherwise,} \end{cases} \forall\, i, j, k. \tag{15}$$

In each time slot, constraint (10) enforces that the sum of time CU $i$ spend on all channels is less than $C_i$. Using discontinuous orthogonal frequency division multiplexing (D-OFDM), the CBS or a CU can aggregate multiple discontinuous orthogonal channels together to form an aggregated channel and then access the aggregated channel [11]–[13]. This way, the CBS or a CU can aggregate and access all the channels assigned to it simultaneously. Therefore, after computing all the $y_{ijk}^h$'s, the CBS transmits to all CUs sequentially according to the $y_{ijk}^h$'s, and each CU is informed of this schedule before transmission so that each CU knows when to access which channel, then it is a feasible and optimal scheduling. Constraint (11) enforces that the transmission time on each channel is within 1 time slot. constraint (12) enforces that the total transmission power of the CBS must not exceed the average power limit $G_{total}$; constraint (13) enforces that there are $\Lambda_j$ CUs to sense each channel $j$; and constraint (15) enforces that the necessary condition for the CBS to transmit to CU $i$ on channel $j$ with power level $k$ is that the resulting SNR must be greater than a predefined threshold $\bar{\gamma}$ such that CU $i$ can successfully decode the received video.

Note that constraint (15) indicates $y_{ijk}^h \ge 0$. Combined with constraint (11), it follows that each $0 \le y_{ijk}^h \le 1$. Therefore, constraint (15) can be rewritten as

$$y_{ijk}^h \begin{cases} \ge 0, & \text{if } G_k d_{ij}/(n_0 B_j) \ge \bar{\gamma} \\ = 0, & \text{otherwise,} \end{cases} \forall\, i, j, k. \tag{16}$$

The upper bound of 1 on the $y_{ijk}^h$'s is thus removed and the problem can be solved more efficiently, since usually LP solvers solve an LP without upper bounds faster than LPs with upper bounds.

## IV. Problem Decomposition

### A. Optimal Sensing Strategy for Problem P0

The formulated problem P0 is an MINLP, which is NP-hard. However, we observe that the optimal sensing strategy can be obtained by solving a relatively easier problem as follows.

We first introduce Lemma 1 as a basis for our later analysis.

*Lemma 1:* The objective value of P0 is a decreasing function of $P_{f_j}$, for all $j = 1, 2, \ldots, N_1$.

*Proof:* Define $F(\mathbf{X}, \mathbf{Y}_h) = \sum_{i=1}^{M}\sum_{j\in\Phi_h}\sum_{k=1}^{K} w_{ijk} \cdot y_{ijk}^h \cdot P(\vec{\mathbf{S}} = \vec{\mathbf{S}}_h)$ and $f(\mathbf{X}, \mathbf{Y}) = \sum_{h=0}^{2^{N_1}-1} F(\mathbf{X}, \mathbf{Y}_h)$. The partial derivative of $F(\mathbf{X}, \mathbf{Y}_h)$ with respect to $P_{f_j}$ is

$$\begin{aligned} \frac{\partial F(\mathbf{X}, \mathbf{Y}_h)}{\partial P_{f_j}} &= -y_{ijk}^h \cdot P(\vec{\mathbf{S}} = \vec{\mathbf{S}}_h) \cdot \sum_{i=1}^{M}\sum_{k=1}^{K}\Big(\beta_i \cdot R_{00}^{ijk} \cdot \\ &((1 - P_{req})P(r_j = 0)P(r_j = 1) + (1 - P_{f_j})P(r_j = 0)^2) + \\ &\alpha_i P(r_j = 0)(R_{00}^{ijk} P_{00}^j + R_{10}^{ijk} P_{10}^j)/B_j\Big) \le 0. \end{aligned}$$

It follows that $\frac{\partial f(\mathbf{X},\mathbf{Y})}{\partial P_{f_j}} = \sum_{h=0}^{2^{N_1}-1} \frac{\partial F(\mathbf{X},\mathbf{Y}_h)}{\partial P_{f_j}} \le 0$. ■

*Theorem 1:* The optimal spectrum sensing strategy to problem P0 can be obtained by solving the following problem SP1.

$$\underline{\mathbf{SP1}} : \ \forall\, j = 1, 2, \ldots, N_1$$

$$\min : \ P_{f_j} = 1 - \prod_{i=1}^{M}(1 - P_{f_{ij}})^{x_{ij}} \tag{17}$$

$$\text{s.t.} \ \sum_{i=1}^{M} x_{ij} = \Lambda_j. \tag{18}$$

*Proof:* Let the optimal solution to the original problem P0 be $(\mathbf{X}', \mathbf{Y}')$ and the solution to SP1 be $\mathbf{X}^*$. Since $\mathbf{X}'$ is optimal to the maximization problem P0, we have $f(\mathbf{X}^*, \mathbf{Y}') \le f(\mathbf{X}', \mathbf{Y}')$.

On the other hand, since $\mathbf{X}^*$ is optimal to the minimization problem SP1, it follows Lemma 1 that $f(\mathbf{X}^*, \mathbf{Y}') \geq f(\mathbf{X}', \mathbf{Y}')$. Therefore we conclude that $f(\mathbf{X}^*, \mathbf{Y}') = f(\mathbf{X}', \mathbf{Y}')$ and $\mathbf{X}^*$ is optimal to problem P0.

After obtaining $\mathbf{X}^*$, we substitute it into problem P0 to solve for $\mathbf{Y}^*$. The $\mathbf{Y}^*$ obtained this way is also optimal to P0, i.e., we have $f(\mathbf{X}^*, \mathbf{Y}^*) \geq f(\mathbf{X}^*, \mathbf{Y}) \geq f(\mathbf{X}, \mathbf{Y})$, for all $\mathbf{X}, \mathbf{Y}$. The proof is completed. ■

From the proof of Theorem 1, we have the following Decomposition Principle for the joint sensing, channel assignment and power allocation problem.

*Corollary 1:* If there is no restriction on the sensing capability for each CU, or $\eta_i \leq \Theta_i$, for all $i$, problem P0 that jointly optimizes spectrum sensing, channel assignment, and power allocation can be decomposed into two sub-problems: one for the optimal spectrum sensing strategy, and the other for the optimal channel assignment and power allocation, without sacrificing optimality.

Problem SP1 can be rewritten as the following problem SP1a,

$$\underline{\textbf{SP1a}}: \quad \forall\, j = 1, 2, \ldots, N_1,$$

$$\max: \quad \sum_{i=1}^{M} x_{ij} \cdot \log_2(1 - P_{f_{ij}}) \tag{19}$$

$$\text{s.t.} \quad \sum_{i=1}^{M} x_{ij} = \Lambda_j, \tag{20}$$

which can be solved easily with an LP solver.

### B. Optimal Solution for a More General Condition

Due to the time constraint (i.e., when the channels are sensed sequentially) or the hardware constraint, it may not be feasible for a CU to sense all the channels. We consider a more general case of spectrum sensing, where each CU $i$ can only sense at most $\Theta_i$ channels simultaneously at a time slot. Therefore the following additional constraint is added to Problem P0. Problem P0 is a special case when $\Theta_i = N_1$.

$$\sum_{j=1}^{N_1} x_{ij} \leq \Theta_i < N_1, \forall\, i. \tag{21}$$

Denote the more general problem as P0a. We still apply the same solution algorithm as in Theorem 1 to Problem P0a, and each channel $j$ will select $\Lambda_j$ users with the best sensing performance (smallest false alarm probability). There are two different cases. First, if a CU $i$ is selected by more than $\Theta_i$ channels, then the new Constraint (21) is violated. Second, if CU $i$ is selected by less than $\Theta_i$ channels, then Constraint (21) is still satisfied and Theorem 1 still holds true and the solution is optimal under the new constraint. Denote $\Pi_j$ as the set of $\Lambda_j$ CUs with the best sensing performance regarding to channel $j$. We can use the following procedure shown in Algorithm 1 to check if each CU $i$ is selected by less than $\Theta_i$ channels.

Algorithm 1 has a polynomial complexity of $\mathcal{O}(MN_1^2)$. We conjecture that if the PUs are widely separated, or when the channels are highly diverse, Theorem 1 will be more likely to hold true under the new practical constraint (21). However, if Theorem 1 is not applicable under (21), we can use a heuristic algorithm to obtain a near-optimal solution to P0a, which is presented in the following section.

**Algorithm 1.** Applicability of Theorem 1

```
for i = 1 : M do
    η_i = 0;
    for j = 1 : N_1 do
        if CU i ∈ Π_j then
            η_i = η_i + 1;
        end
    end
    if η_i > Θ_i then
        Theorem 1 is not applicable;
        Break;
    end
end
```

**Algorithm 2.** Heuristic Spectrum Sensing Algorithm

```
1  Sort the N_1 channels in descending order of P(r_j = 0) and
   let the sorted channel set be Ξ;
2  for j = 1 : N_1 do
3      Let j' = Ξ(j);
4      Solve problem SP1b and denote the solution as Θ_j' (underlined);
5      if Θ_j' (underlined) = ∅ then
6          Channel j' is determined to be busy;
7      end
8      for i = 1 : M do
9          if x_ij' = 1 then
10             Θ_i = Θ_i − 1;
11         end
12     end
13 end
```

### C. Heuristic Spectrum Sensing Algorithm

The idea of heuristic algorithm to Problem P0a is to sort the $N_1$ channels according to $P(r_j = 0)$, for all $j = 1, 2, \ldots, N_1$, in the descending order, and then minimize $P_{f_j}$, for all $j = 1, \ldots, N_1$, sequentially. The heuristic spectrum sensing algorithm is presented in Algorithm 2. In Line 4, the following problem SP1b is solved.

$$\underline{\textbf{SP1b}}: \quad \forall\, j = 1, 2, \ldots, N_1,$$

$$\min: P_{f_{j'}} = 1 - \prod_{i=1}^{M} (1 - P_{f_{ij'}})^{x_{ij'}} \tag{22}$$

$$\text{s.t.} \quad \sum_{i=1}^{M} x_{ij'} = \Lambda_{j'} \tag{23}$$

$$x_{ij'} \leq \Theta_i. \tag{24}$$

In Lines 5∼7, if there is no feasible solution to problem SP1b, there is not a sufficient number of CUs to sense channel $j'$, and we conservatively assume that channel $j'$ is busy to avoid collision with PUs. Each time if CU $i$ is assigned to sense a channel, $\Theta_i$ is decreased by 1 as in Lines 9∼11. When $\Theta_i$ reaches 0, constraint (24) will prevent CU $i$ to be assigned to sense any more channels.

In Section VII, we will show that the performance of Algorithm 2 is very close to that of the optimal sensing strategy in terms of both sensing performance and the expected overall QoS, even when $\Theta_i \ll N_1$.

### D. Optimal Chanel Assignment and Power Allocation Solution

After obtaining $\mathbf{X}^*$, cooperative sensing is conducted and the CBS determines the set of available channels based on sensing results, as shown in Fig. 1. From now on, we omit the subscript (or superscript) $h$ in all the symbols, since the cooperative sensing results on the $N_1$ channels is already determined. Denote the number of channels sensed idle as $N_2$, and re-index the $N_2$ idle channels as $1, 2, \ldots, N_2$. Then the remaining channel assignment and power allocation problem SP2 can be written as follows.

$$\underline{\mathbf{SP2}} : \max : \quad \sum_{i=1}^{M}\sum_{j=1}^{N_2}\sum_{k=1}^{K} w_{ijk} \cdot y_{ijk} \tag{25}$$

$$\text{s.t.} \quad \sum_{j=1}^{N_2}\sum_{k=1}^{K} y_{ijk} \le C_i, \forall\, i \tag{26}$$

$$\sum_{i=1}^{M}\sum_{k=1}^{K} y_{ijk} \le 1, \forall\, j \tag{27}$$

$$\sum_{i=1}^{M}\sum_{j=1}^{N_2}\sum_{k=1}^{K} y_{ijk} \cdot G_k \le G_{total}$$

$$\text{Constraint (16).} \tag{28}$$

In practice, there may be a large number of CUs and licensed channels, and the CBS also has a great flexibility to choose the power level for transmission on a channel. Therefore, the constraint matrix of SP2 could be huge and it may be hard to solve with an LP solver due to its size. In the next section, we propose to use the Column Generation (CG) method [15] to solve SP2 and derive a decentralized algorithm for better scalability. With the proposed CG method, the CBS and CUs solve different sub-problems, thus alleviating the computational burden on the CBS.

## V. Column Generation Method to Solve SP2

### A. Dantzig-Wolfe Decomposition

We first reformulate problem SP2 from the standard form into a *disaggregated formulation* by applying Dantzig-Wolfe decomposition of LP problems [22].

For $i = 1, 2, \ldots, M$, let $\overline{\Theta}_i = \left\{\chi_i^1, \chi_i^2, \ldots, \chi_i^{Q_i}\right\}$ denote the set of feasible channel assignment and power allocation schemes to CU $i$. Then

$$\chi_i^q = \left\{y_{ijk}^q, j = 1, 2, \ldots, N_2, k = 1, 2, \ldots, K\right\},$$

for $q = 1, 2, \ldots, Q_i$, is a feasible scheme satisfying all the constraints, where $y_{ijk}^q = 1$ if the CBS transmits to CU $i$ on channel $j$ at power level $k$, and $y_{ijk}^q = 0$ otherwise. Thus, the feasible schemes are indeed the extreme points of the the feasible region of SP2, which is the key for Dantzig-Wolfe decomposition [22].

Introduce a variable $0 \le z_i^q \le 1$ to denote the amount of time the CBS transmits using feasible scheme $\chi_i^q$ within a time slot. Let the "utility" gained by using $\chi_i^q$ for CU $i$ as $\varpi_i^q = \sum_{j=1}^{N_2}\sum_{k=1}^{K} w_{ijk} \cdot y_{ijk}^q$. Then SP2 can be represented in a *set-partition form*, termed the Master Problem (MP), as

$$\underline{\mathbf{MP}} : \max : \quad \sum_{i=1}^{M}\sum_{q=1}^{Q_i} \varpi_i^q \cdot z_i^q \tag{29}$$

$$\text{s.t.} \quad \sum_{q=1}^{Q_i} z_i^q \le 1, \forall\, i, \tag{30}$$

$$\sum_{i=1}^{M}\sum_{q=1}^{Q_i}\left(\sum_{k=1}^{K} y_{ijk}^q\right) z_i^q \le 1, \forall\, j, \tag{31}$$

$$\sum_{i=1}^{M}\sum_{q=1}^{Q_i}\left(\sum_{j=1}^{N_2}\sum_{k=1}^{K} y_{ijk}^q \cdot G_k\right) z_i^q \le G_{total}, \tag{32}$$

$$0 \le z_i^q \le 1, \forall\, i, q. \tag{33}$$

Constraint (30) ensures that $0 \le y_{ijk} = \sum_{q=1}^{Q_i} y_{ijk}^q z_i^q \le 1$, for all $i, j, k$; constraints (31) and (32) correspond to constraints (27) and (28), respectively; and constraints (26) and (16) are specified in the INitialization Problem (INP) and Pricing Problem (PP) defined next in Section V-B. For convenience of our later discussion, the problem containing a subset of the columns and cost coefficients (variables) of the MP is called Restricted MP (RMP).

### B. Design of the Column Generation Method

Obviously, it is infeasible to solve the MP directly due to the exponential number of columns. However, usually most of the variables in the optimal solution to the MP are equal to zero, with only a small number of positive-valued variables. The MP solution can be re-optimized iteratively by finding the variables having the potential to improve the objective value at each iteration. This is done by iteratively solving the PP, which examines whether there exists a variable with a negative (in the case of a minimization problem) or positive (in the case of a maximization problem) reduced cost, and then generates the corresponding column to add it to the RMP.

The RMP contains only a small subset of all the feasible columns and variables of the MP and thus can be solved quickly. The simplex multiplier obtained from the RMP will be passed to the PP to identify a new column to enter the RMP again, until there is no variables whose reduced cost is negative (in the case of a minimization problem) or positive (in the case of a maximization problem). Thus an optimal feasible solution to the MP is found. The purpose of RMP is to generate the simplex multiplier for solving the PP.

The CG based Distributed Optimization Algorithm (CDOA) includes the following six steps.

**Algorithm 2.** CG Based Distributed Optimization Algorithm

**Step 1**: CU $i$ solves the following $i$-th INP and reports its solution to the CBS, $i = 1, 2, \ldots, M$.

$$\underline{\textbf{INP}} : \max : \sum_{j=1}^{N_2}\sum_{k=1}^{K} w_{ijk} \cdot y_{ijk} \tag{34}$$

$$\text{s.t.} \sum_{j=1}^{N_2}\sum_{k=1}^{K} y_{ijk} \le C_i, \tag{35}$$

$$\sum_{k=1}^{K} y_{ijk} \le 1, \forall\, j, \tag{36}$$

$$y_{ijk} \begin{cases} \in \{0,1\}, & \text{if } G_k d_{ij}/(n_0 B_j) \ge \bar{\gamma} \\ = 0, & \text{otherwise,} \end{cases} \forall\, j, k. \tag{37}$$

Each of the $M$ solutions generates a feasible column of the MP. The CBS uses the $M$ feasible columns and the corresponding cost coefficients to initiate the RMP, which has the same formulation with the MP, but with $Q_i = 1$, for all $i = 1, 2, \ldots, M$.

**Step 2**: The CBS solves the RMP, from which a vector of simplex multiplier $\Omega^T = (\nu^T, \mu^T, \varphi)$ is obtained, where $(\cdot)^T$ denotes the transpose of a vector, $\nu^T$ is a $1 \times M$ vector with the $i$-th entry $\nu_i$ corresponding to the $i$-th constraint in the RMP, $\mu^T$ is a $1 \times N_2$ vector with the $j$-th entry $\mu_j$ corresponding to the $(M + j)$-th constraint in the RMP, and $\varphi$ is the simplex multiplier corresponding to the last constraint in the RMP. The objective value of the RMP is a *lower bound* to the MP.

**Step 3**: The CBS broadcasts $\Omega^T$ to all CUs and assigns CU $i$ to solve the following $i$-th PP, to find the column and the corresponding variable with the most positive reduced cost [15] to enter the RMP to improve the objective value of the MP.

$$\underline{\textbf{PP}} : \max : \quad \Delta_i = \sum_{j=1}^{N_2}\sum_{k=1}^{K} (w_{ijk} - \mu^j - \varphi G_k) \cdot y_{ijk} - \nu_i \tag{38}$$

$$\text{s.t.} \quad \sum_{j=1}^{N_2}\sum_{k=1}^{K} y_{ijk} \le C_i, \tag{39}$$

$$\sum_{k=1}^{K} y_{ijk} \le 1, \forall\, j,$$
$$\text{Constraint (37).} \tag{40}$$

**Step 4**: Each CU $i$ decides when to report its optimal solution to the $i$-th PP to the CBS according to a delay

$$\tau_i = \xi(\Delta_i), \tag{41}$$

where $\xi(\cdot)$ denotes a decreasing function of $\Delta_i$. Define an index $a = \arg\max_{i=1,\ldots,M}\{\Delta_i\}$. In case that $\Delta_a > 0$, then the current optimal solution to the RMP is not optimal to the MP and CU $a$ sends its solution in the earliest time $\tau_a$ (since it has the maximum value $\Delta_a$). Other CUs overhearing CU $a$'s message will not send their respective messages. In case that $\Delta_a \le 0$, the current optimal solution to the RMP is also optimal to the MP, and no CU sends message to CBS.

**Step 5**: The CBS verifies the optimality of the current solution: if nothing is received from the CUs after a predefined period of time, the CBS concludes that $\Delta_a \le 0$ and thus the CG method is terminated; otherwise, go to Step 6.

**Step 6**: For index $a = \arg\max_{i=1,\ldots,M}\{\Delta_i\}$, let $Q_a = Q_a + 1$ and generate the column

$$H_a^{Q_a} = \left[ e_a, \sum_{k=1}^{K} y_{a1k}^{Q_a}, \ldots, \sum_{k=1}^{K} y_{aN_2k}^{Q_a}, \sum_{j=1}^{N_2}\sum_{k=1}^{K} y_{ajk}^{Q_a} \cdot G_k \right]^T \tag{42}$$

with the solution to the $a$-th PP derived in Step 3, where $e_a$ is a $1 \times M$ unit vector with the $a$-th entry being 1. Add the column and the corresponding variable $z_a^{Q_a}$ to the RMP and go to Step 2.

## VI. Upper Bound, Complexity and Time Efficiency

### A. Upper Bound for the MP

In the following, we derive an upper bound for the optimal objective value of the MP in each iteration of the CG method.

*Theorem 2:* At each iteration, let $\Omega^T$ be the simplex multiplier vector of the RMP; $\Delta_a = \max_{i=1,\ldots,M}\{\Delta_i\}$ be the most positive reduced cost obtained from the PPs; $\vec{b}$ be a $(M + N_2 + 1) \times 1$ column vector with the $i$-th entry being the value of right hand side of the $i$-th constraint of the RMP, $i = 1, 2, \ldots, (M + N_2 + 1)$; $\vec{g}$ be a $(M + N_2 + 1) \times 1$ column vector as $g = (\underbrace{11\ldots1}_{M\ ones}\ \underbrace{00\ldots0}_{(N_2+1)\ zeros})^T$. Then an upper bound for the MP can be derived as: $\overline{\Omega}^T\vec{b} = (\Omega^T + \Delta_a\vec{g})\vec{b}$.

*Proof:* Let $\widehat{\Omega}^T$ be a feasible solution to the dual problem of the MP (termed DMP), according to the relationship between the dual and primal problems [15], we have

$$\widehat{\Omega}^T H_i^q \ge \varpi_i^q,\ i = 1, \ldots, M,\ q = 1, \ldots, Q_i, \tag{43}$$

where $H_i^q$ is given in (42). As discussed, at each iteration we can obtain a simplex multiplier vector $\Omega^T$ by solving the RMP, as well as the most positive reduced cost by solving the PP.

$$\Delta_a = \max_{i,q}\{\varpi_i^q - \Omega^T H_i^q\}$$
$$\Rightarrow \Omega^T H_i^q \ge \varpi_i^q - \Delta_a, i = 1, \ldots, M, q = 1, \ldots, Q_i, \tag{44}$$

where $\Delta_a > 0$. Denote $\overline{\Omega}^T = (\Omega^T + \Delta_a\vec{g})$ and multiply its both sides by $H_i^q$. We have

$$\overline{\Omega}^T H_i^q = \Omega^T H_i^q + \Delta_a\vec{g}H_i^q \Rightarrow \overline{\Omega}^T H_i^q = \Omega^T H_i^q + \Delta_a \cdot 1$$
$$\Rightarrow \overline{\Omega}^T H_i^q - \Delta_a = \Omega^T H_i^q \Rightarrow \overline{\Omega}^T H_i^q - \Delta_a \ge \varpi_i^q - \Delta_a$$
$$\Rightarrow \overline{\Omega}^T H_i^q \ge \varpi_i^q,\ i = 1, \ldots, M,\ q = 1, \ldots, Q_i.$$

The first inequality is from (44). This means that $\overline{\Omega}^T$ is a feasible solution to the DMP. By duality, the corresponding dual LP of a maximization LP is a minimization LP [15]. So the DMP is a minimization LP, and $\overline{\Omega}^T$ is a feasible solution to the DMP.

Let the optimal solution to DMP be $\underline{\Omega}^T$. It follows that

$$\overline{\Omega}^T \vec{b} \geq \underline{\Omega}^T \vec{b} = \Upsilon^*, \tag{45}$$

where $\Upsilon^*$ is the optimal objective value of the minimization LP DMP. Due to *Strong Duality*, $\Upsilon^*$ is also the optimal objective value of the MP. It follows that $\overline{\Omega}^T \vec{b} = (\Omega^T + \Delta_a \vec{g})\vec{b}$ is an upper bound for the MP according to (45). ■

### B. Complexity and Optimality Analysis

In the general problems solved by the CG method, the INP and PP problems are at least as hard as the one dimensional 0-1 Knapsack problem, which is NP-hard [25].

However, an interesting characteristic of the INP and PP in our case is that the coefficients of the constraint matrix in the INP and PP are either 0 or 1, such that the *unimodularity property* [24] is satisfied in both problems. As a result, both the INP and PP have the optimal solution with their LP relaxations, and thus they can be solved with the Simplex method [15], [25]. Again, the upper bound of 1 on $y_{ijk}$ can be removed as in P0.

*Lemma 2:* The INP and PP are indeed LPs and thus can be solved with the Simplex method with a polynomial-time average-case complexity.

Although the INP and PP can be solved with the Simplex method which has a average complexity of polynomial-time, it is possible that the Simplex method will require exponential time in extreme cases. Here we introduce a Greedy algorithm which solves the INP and PP in strongly polynomial time and still gets the optimal solution, to reduce the time complexity of CDOA.

The basic idea of our Greedy Algorithm is that, for the INP or PP of each CU, the decision variables corresponding to the combinations of channel and power level that having the highest "utilities" among all possible combinations and still satisfy the constraints of the INP or PP, take value of 1, while the other decision variables take value of 0. In this way a feasible and optimal solution is guaranteed.

Note that Line 1 to Line 5 of Greedy Algorithm executes in Step 1 of CDOA to solve the INP and thus executes only once, in order to further reduce the computation complexity of our Greedy Algorithm, while Line 6 to Line 36 executes in Step 3 of CDOA to solve the PP and thus may executes multiple times during a whole operation of CDOA.

*Theorem 3:* The Greedy Algorithm yields the optimal solution for the INP and PP.

*Proof:* We can see from constraint (40) that each channel can be chosen at most once with one power level on it, and for each combination of channel and power level, there is an associated utility $\xi_{ijk}$.

Therefore, in order to maximize the total utility:

1) For each channel, we choose the combination of channel and power level having the greatest utility among all the $K$ combinations;
2) If a combination has a negative utility, say, $\xi_{ijk}$, then the combination should not be used, i.e., $y_{ijk} = 0$ in the optimal solution. So the chosen combination should have a positive utility;
3) From constraint (37) we know that if the resulting SNR of a combination, say $\delta_{ijk}$, is less than the threshold, then the combination should not be used. So the chosen combination should have a resulting SNR greater than the threshold.

We form a set using the combinations where each combination must satisfy the above three conditions.

From constraint (39) we know that each CU can use at most $C_i$ channels. Therefor, from the above set, we choose $\rho_i$ combinations having the greatest utilities among all the combinations, where $\rho_i = C_i$ if the number of elements in this set is greater or equal to $C_i$, and $\rho_i < C_i$ otherwise.

In this algorithm, all the constraints are used and thus the solution is feasible, and from the discussion we know that the solution is also optimal. ■

For the greedy algorithm that solves the INP at optimality, we just need to replace $\xi_{ijk} = w_{ijk} - \mu^j - \varphi G_k$ with $\xi_{ijk} = w_{ijk}$ at Line 11 of the above algorithm.

*Lemma 3:* The Greedy Algorithm has a polynomial computational complexity of $\mathcal{O}(\max\{N_2^2, N_2 K\})$.

The proof is omitted for brevity.

### C. Time Efficiency of CDOA

Finally we analyze the time efficiency of CDOA. We compare the time needed to solve the MP with CODA (with a distributed parallel execution), $t_1$, and that without using CODA (with a centralized sequential execution), $t_2$.

Let the number of iterations of CDOA be $L$. Let $\tau_{ml}$ be the amount of time to process the RMP, $\tau_{pl}$ the amount of time to process a PP, $\tau_b$ the time for the CBS to broadcast the simplex multiplier to the CUs, $\tau_{rl}$ the time for CU $a$ (where $a = \max_{i=1,\ldots,M}\{\Delta_i\}$) to report its solution to the $a$-th PP to the CBS, all in the $l$-th iteration of CDOA. Also assume that the time to broadcast the simplex multiplier to the CUs by the CBS is negligible. The *time saving* $s = t_2 - t_1$ achieved by the distributed, parallel execution can be approximated as

$$\begin{aligned} s &= \sum_{l=1}^{L}(\tau_{ml} + M \cdot \tau_{pl}) - \sum_{l=1}^{L}(\tau_{ml} + \tau_{pl} + \tau_b + \tau_{rl}) \\ &= \sum_{l=1}^{L}((M-1)\cdot\tau_{pl} - \tau_b - \tau_{rl}). \end{aligned} \tag{46}$$

Since the size of the PP does not change during each iteration, we assume that $\tau_{pl} = \tau_p$. Then (46) can be rewritten as

$$s = \tau_p \sum_{l=1}^{L}\left((M-1) - \frac{\tau_b + \tau_{rl}}{\tau_p}\right). \tag{47}$$

In (47), $\tau_b$ and $\tau_{rl}$ are usually negligible compared with $\tau_p$ for even a modest problem having hundreds or thousands of constraints and variables. When the number of simplex iterations is proportional to the number of constraints, the overall

TABLE I
SIMULATION PARAMETERS

| *Parameter* | *Value* | *Parameter* | *Value* |
|---|---|---|---|
| $M$ | 30 | $\bar{\gamma}$ | $-25$ dB |
| $N_1$ | 30 | $\Lambda_j$ | 3 |
| $K$ | 10 | $r_p$ | 50 |
| $\kappa$ | $10^4$ | $\Theta_i$ | 3 |
| $d_{ij}$ | $-15 \sim -9$ dB | $\gamma_{ij}$ | $-100 \sim 0$ dB |
| $C_i$ | 3 | $G_{total}$ | 50 |
| $G_k$ | $10^{-\frac{k}{10}}$ | $P_{req}$ | 0.99 |
| $f_s$ | $10^6\ H_z$ | $\min_j\{P(r_j\!=\!0)\}$ | 0.2 |
| $n_0$ | $10^{-6}$ | $\max_j\{P(r_j\!=\!0)\}$ | 0.9 |
| $B_j$ | $10^6\ H_z$ | | |

cost of the simplex method is $\mathcal{O}(m^4 + nm^2)$ arithmetic operations for a problem having $m$ constraints and $n$ variables [25]. Therefore, $s$ can be further approximated by

$$s = \mathcal{O}\left(LM(N_2^4 + KN_2^3)\right). \quad (48)$$

It can be seen that $s$ is an increasing function of $L$, $M$, $N_2$, and $K$, which represent the size of the problem. Such improvement in time efficiency demonstrate the advantages of the distributed CG-based algorithm.

## VII. SIMULATION RESULTS AND ANALYSIS

In this section, Matlab simulation results are used to demonstrate the performance of the proposed algorithms. Unless specified, the value of simulation parameters are as shown in Table I. Each simulated point in the figures is obtained by repeating the simulation $r_p = 50$ times with different random seeds, and 95% confidence intervals are computed and plotted in the figures to guarantee credible results. Three widely used videos in the uncompressed YUV4MPEG format are used in the simulation. They represent three levels of motions: slow (Suzie), medium (Carphone), and quick (Football) [28], and the accuracy of the QoS model has been tested in our prior work [29]. In the figure, the markers are obtained by experiment with the real video sequences, and the lines are obtained by linear regression. Video QoS parameters $\alpha_i$ and $\beta_i$ for the three sequences are calculated based on their respective linear regression plots.

In Figs. 2 and 3 we compare the performance of the optimal sensing strategy and that of the Heuristic algorithm, in terms of sensing performance and the resulting overall Y-PSNR of the received videos, respectively, We consider four cases that the number of sensors a CU has in the Heuristic algorithm $\Theta_i$ = 3, 4, 5, and 6, where $\Theta_i \ll N_1 = 30$. Note that $M \cdot \Theta_i \geq N_1 \cdot \Lambda_j$ is a necessary condition to have any channel sensed by at lest $\Lambda_j$ sensors. In Fig. 2, the legend 'Idle channels' means the number of channels cooperatively sensed idle; 'Missed channels' means the number of channels cooperatively sensed idle while these channels are actually busy, so the number of channels that are sensed idle and are actually idle is the difference between the two.

From these two figures we can see that the heuristic algorithm achieves a performance close to the optimal sensing strategy when $\Theta_i \geq 3$. In this case, we have $M \cdot \Theta_i = N_1 \cdot \Lambda_j$.

**Algorithm 3.** Greedy Algorithm

**for** $j = 1 : N_2$ **do**
  **for** $k = 1 : k$ **do**
    Compute $\delta_{ijk} = G_k d_{ij}/(n_0 B_j)$;
  **end**
**end**
**for** $j = 1 : N_2$ **do**
  **for** $k = 1 : K$ **do**
    **if** $\delta_{ijk} < \bar{\gamma}$ **then**
      Set $\xi_{ijk} = \sigma$, where $\sigma < 0$ and is a fixed constant;
      **else**
        Compute $\xi_{ijk} = w_{ijk} - \mu^j - \varphi G_k$;
      **end**
    **end**
  **end**
**end**
**for** $j = 1 : N_2$ **do**
  $\hat{\xi}_{ij} = \arg\max_{k=1,\ldots,K}\{\xi_{ijk}\}$;
**end**
Let $\rho_i = 0$;
**for** $j = 1 : N_2$ **do**
  **if** $\widehat{\xi}_{ij} > 0$ **then**
    $\rho_i = \rho_i + 1$;
  **end**
**end**
Sort the $\rho_i$ channels in the decreasing order of $\widehat{\xi}_{ij}$ and let the sorted channel set be $\mathcal{A}$;
**if** $\rho_i \leq C_i$ **then**
  **for** $j = 1 : \rho_i$ **do**
    Set $y_{i\mathcal{A}(j)k'} = 1$;
  **end**
  **else**
    **for** $j = 1 : C_i$ **do**
      Set $y_{i\mathcal{A}(j)k'} = 1$;
    **end**
  **end**
**end**

A channel $j$, may not have 'good' CUs (i.e., with a relative small $P_{f_{ij}}$) to sense it, since these CUs may have already been assigned to another channel $j'$, which has a higher $P(r_{j'} = 0)$ and thus has a higher priority of being optimized. Thus channel $j$ may be discarded due to insufficient CUs to sense it, resulting in a lower value of SP2 due to fewer available channels for video streaming. Recall that $\Theta$ is the number of channels that a CU can sense at a time slot. Once a CU is assigned to sense $\Theta$ channels, it is deleted from the remaining CUs list. Therefore, as $\Theta$ is increased, which means that a CU can sense more channels at a time slot, it is more likely that each channel can be sensed by the best CUs. Thus the sensing performance of each channel will be improved. We found that when $\Theta$ is selected from 3 to 6, the result is sufficiently good. However, as the last three sub-figures shows, the Heuristic algorithm achieves almost the same performance as the optimal sensing strategy when $\Theta_i$ = 4, 5, and 6. Thus even the channels having a lower

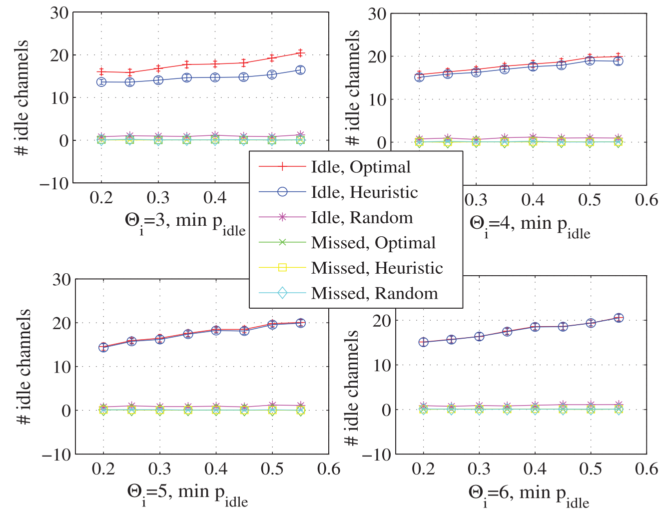


Fig. 2. Optimal versus Heuristic Sensing in terms of sensing performance for $\Theta = 3, 4, 5, 6$.

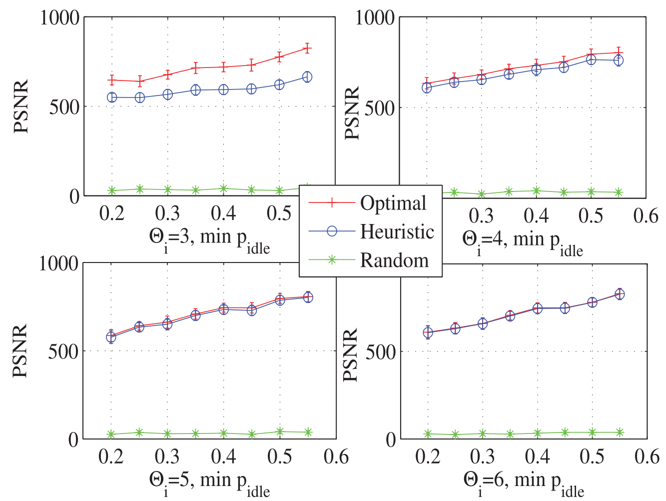


Fig. 3. Optimal versus Heuristic Sensing in terms of overall Y-PSNR in dB for $\Theta = 3, 4, 5, 6$.

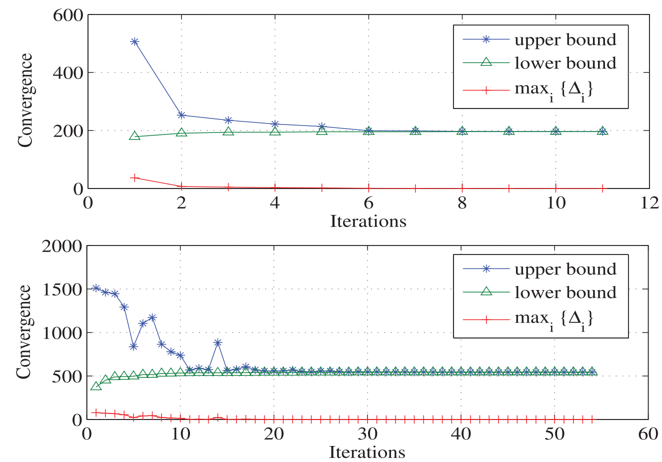


Fig. 4. Convergence performance of CG: (Upper) $M = 9, N_1 = 18$ and (Lower) $M = 15, N_1 = 30$. $C_i = 2$, for all $i = 1, 2, \ldots, M$.

priority will have a higher chance to be sensed by 'good' CUs. Then $P_{f_j}$, $j = 1, 2, \ldots, N_2$ is more likely to be reduced, and the objective function value is improved. Note that both the heuristic and the optimal sensing have the same computational complexity and workload for the 'good' cognitive users, since in both algorithms each channel will be sensed by the CUs with the best sensing performance. The only difference is that in the optimal sensing strategy, the sensing of each channel is optimized at the same time; while in the heuristic algorithm, the sensing of each channel is optimized sequentially, in the order of decreasing priority, which is determined by its probability of being idle.

Besides, we also compare the sensing performance and overall Y-PSNR performance of our proposed Heuristic algorithm with a Benchmark Algorithm called the Random Algorithm, as the authors in [12] did, which randomly assigns $\Lambda_j$ CUs to sense channel $j$, $\forall j$. We can see that the Heuristic algorithm outperforms the Random Algorithm significantly, the main reason for which is that the Random Algorithm doesn't consider the sensing accuracy of different CUs for different channels. It is very likely that for a particular channel, the Random Algorithm will assign CUs having a high false alarm probability to this channel to sense it, which results in a great probability of this channel being false alarmed. That's why the number of idle channels found by the Random Algorithm is very small, although the number of missed detected channels of the Random Algorithm is close with that of the Heuristic Algorithm and the optimal sensing strategy. Under such sensing performance, it is within our expectation that the overall Y-PSNR performance of the Random Algorithm falls behind that of the Heuristic Algorithm drastically. Therefore, we claim that Heuristic Algorithm provides a much better performance than Random Algorithm does.

Fig. 4 demonstrates the convergence of the CG method in two cases: $M = 9$, $N_1 = 18$ (the upper figure) and $M = 15$, $N_1 = 30$ (the lower figure). We set $C_i = 2$, for all $i = 1, 2, \ldots, M$ in both cases. We have the following observations. (i) The number of iterations is positively correlated to the problem size, since as the number of CUs and channels grows, there may be more feasible schemes to improves the current objective value at a specific iteration. (ii) The most positive reduced cost $\max_i \{\Delta_i\}$ tends to decrease over iterations. This trend is the result of the greedy approach of the CG algorithm, which means that the algorithm chooses the feasible scheme having the most positive reduced cost (thus possibly having the greatest potential to improve the current objective value of the MP) from the remaining candidate feasible schemes, to enter the RMP at each iteration. (iii) The increment of the objective function at a specific iteration is positive correlated to $\max_i \{\Delta_i\}$. This follows directly from the above discussions. (iv) The upper bound to the optimal objective function value converges quickly, and is also positively correlated to $\max_i \{\Delta_i\}$. From Theorem 2, the upper bound at a certain iteration $\overline{\Omega}^T b$ is actually a positive function of $\max_i \{\Delta_i\}$ ($\Delta_a$ in Theorem 2) at this iteration. Since $\max_i \{\Delta_i\}$ drops quickly and become very close to 0, $\overline{\Omega}^T b$ drops and then converges to the optimal objective function value quickly. Note that $\overline{\Omega}^T b$ is not necessary decreasing as the iteration goes. There are two main reasons: (a) $\max_i \{\Delta_i\}$ is

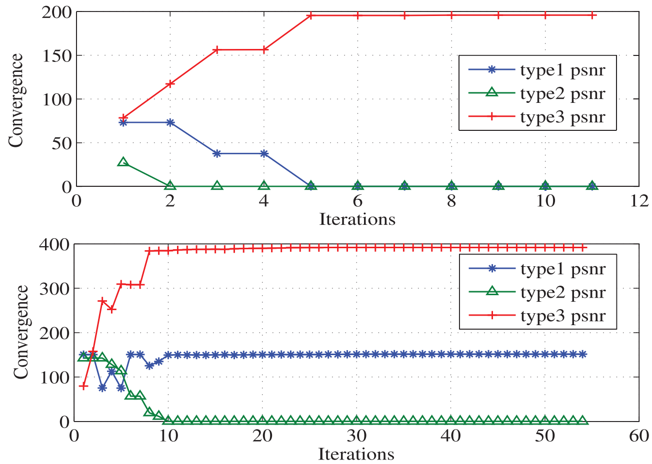


Fig. 5. PSNR comparison for three types of videos achieved by CG: (Upper) $M = 9$, $N_1 = 18$ and (Lower) $M = 15$, $N_1 = 30$. $C_i = 2$, for all $i = 1, 2, \ldots, M$.

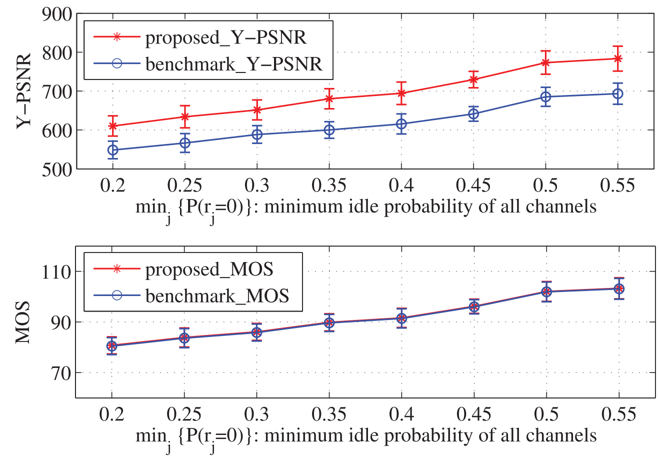


Fig. 6. QoS and QoE comparison between the proposed scheme and the benchmark: (Upper) QoS comparison and (Lower) QoE comparison.

not necessary decreasing as the iteration goes although it shows a trend of decreasing. (b) $\overline{\Omega}^T b$ also depends on the simplex multiplier $\Omega_T$, whose convergence is hard to analysis.

The convergence of the Y-PSNR of each type of received video is shown in Fig. 5. It is observed that for the type 3 video, the Y-PSNR shows a tendency of increasing as the iteration goes. For the type 1 video, the Y-PSNR shows a trend of decreasing, especially in the lower plot in Fig. 5. The Y-PSNR of type 2 video is decreased in the upper plot and doesn't show much change in the lower plot. However, the overall Y-PSNR of the three types of videos increases along with the iterations. The main reason is that to improve the overall Y-PSNR, at each iteration, the CG algorithm preferentially assigns channel and power resources to the type of video which may have the greatest potential for improving the overall Y-PSNR. Recall that under the same data rate, the Y-PSNR of type 3 video is the greatest and the Y-PSNR of type 1 video is the smallest among the three types. Thus on the condition that the number of available channels is limited and is less than the number of video sessions, channel and power resources will be taken from what have been allocated to type 1 videos and then be assigned to type 3 videos first, then to type 2 video. Besides, the magnitude of changes of the three types of videos decreases quickly over iterations, which is in consistence with the change of the overall Y-PSNR, as shown in Fig. 4.

In Fig. 6, we compare the proposed algorithm with a benchmark scheme [26], with respect to both QoS (i.e., Y-PSNR) and QoE of the received videos. In the benchmark scheme, an idle channel, say channel $j$, with a higher $P(r_j = 0)$ is assigned to a CU, say CU $i$, that is less delay-tolerant and thus have a higher priority to use the available channels. In our simulation we randomly assign an integer priority level ranging from 1 to 3 to each CU. We use the MOS (Mean Opinion Score) to evaluate QoE and the MOS model is adopted from [27] as the benchmark scheme does [26].

It can be seen that when $\min_j\{P(r_j = 0)\} \geq 0.35$, the Y-PSNR achieved by the benchmark scheme is about 100 dB lower than that achieved by the proposed scheme, which means that each CU with our proposed scheme has an average Y-PSNR of 3 dB higher than that with the benchmark scheme. Both curves tends to increase as $\min_j \left\{P(r_j = 0)\right\}$ increases, since more channels are likely to be available for video transmission. Within the entire range of $\min_j \left\{P(r_j = 0)\right\}$, our proposed algorithm outperforms the benchmark scheme in terms of Y-PSNR, because the solution to the problem in [26] is not necessary optimal to our problem and thus yields a lower bound to our problem. Furthermore, we find that the overall MOS of the two schemes are almost the same. Thus our proposed algorithm achieves a considerably higher QoS and a comparable QoE performance as the benchmark scheme.

## VIII. Conclusion

In this paper, we investigated the problem of QoS-driven multi-user video streaming over cellular CRNs. We showed that there exists a *decomposition principle* in the optimal joint design of spectrum sensing, channel assignment, and power allocation that circumvents the *curse of dimensionality* in general MINLPs. The decomposed spectrum sensing problem was solved with an optimal algorithm, along with a heuristic algorithm that is much less demanding on CUs' hardware. A CG-based decentralized channel assignment and power allocation algorithm was next developed to relieve the computation burden on the CBS. And a Greedy Algorithm which solves the sub-problems generated during the CG based algorithm at optimality in polynomial time is proposed to reduce the computational complexity. We analyzed the complexity and time efficiency, and derive an upper bound for the CG-based algorithm, and validated its performance with simulations.

## References


[1] Cisco. (2014, Feb.). *Visual Networking Index (VNI)* [Online]. Available: http://www.cisco.com/

[2] Z. He and S. Mao, "QoS driven multi-user video streaming in cellular CRNs: The case of multiple channel access," in *Proc. IEEE Int. Mobile Ad Hoc Sensor Syst.*, Philadelphia, PA, USA, Oct. 2014, pp. 28–36.

[3] Y. Zhao, S. Mao, J. Neel, and J. H. Reed, "Performance evaluation of cognitive radios: Metrics, utility functions, and methodologies," *Proc. IEEE*, vol. 97, no. 4, pp. 642–659, Apr. 2009.

[4] D. Hu, S. Mao, Y. T. Hou, and J. H. Reed, "Scalable video multicast in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 3, pp. 334–344, Apr. 2010.
[5] D. Hu and S. Mao, "Streaming scalable videos over multi-hop cognitive radio networks," *IEEE Trans. Wireless. Commun.*, vol. 11, no. 9, pp. 3501–3511, Nov. 2011.
[6] Y. Xu, D. Hu, and S. Mao, "Relay-assisted multi-user video streaming in cognitive radio networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 10, pp. 1758–1770, Oct. 2014.
[7] S. Soltani and M. Mutka, "Decision tree modeling for video routing in cognitive radio mesh networks," *in Proc. IEEE Int. Symp. Workshops World Wireless Mobile Multimedia Netw.*, Madrid, Spain, Jun. 2013, pp. 1–9.
[8] S. Li, T. Luan, and X. Shen, "Channel allocation for smooth video delivery over cognitive radio networks," *in Proc. IEEE Global Telecommun. Conf.*, Miami, FL, USA, Dec. 2010, pp. 1–5.
[9] R. Yao, Y. Liu, J. Liu, P. Zhao, and S. Ci, "Perceptual experience oriented transmission scheduling for scalable video streaming over cognitive radio networks," *in Proc. IEEE Global Telecommun. Conf.*, Atlanta, GA, USA, Dec. 2013, pp. 1681–1686.
[10] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2053–2071, May 2008.
[11] X. Zhang and H. Su, "CREAM-MAC: Cognitive radio-enabled multi-channel MAC protocol over dynamic spectrum access networks," *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 9, pp. 110–123, Feb. 2011.
[12] H. Su and X. Zhang, "Cross-layer based opportunistic MAC protocols for QoS provisionings over cognitive radio wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 118–129, Jan. 2008.
[13] H. A. B Salameh, M. Krunz, and D. Manzi, "Spectrum bonding and aggregation with guard-band awareness in cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 13, no. 3, pp. 569–581, Mar. 2014.
[14] M. Wien, H. Schwarz, and T. Oelbaum, "Performance analysis of SVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 1194–1203, Sep. 2007.
[15] M. S. Bazaraa, J. J. Jarvis, and H. D. Sherali, *Linear Programming and Network Flows*, 4th ed. Hoboken, NJ, USA: Wiley, 2009.
[16] Y. C. Liang, Y. Zheng, E. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless. Commun.*, vol. 7, no. 4, pp. 1326–1337, Apr. 2008.
[17] W. Wang, B. Kasiri, J. Cai, and A. S. Alfa, "Channel assignment of cooperative spectrum sensing in multi-channel cognitive radio networks," *in Proc. IEEE Int. Conf. Commun.*, Kyoto, Japan, Jun. 2011, pp. 1–5.
[18] A. W. Min and K. G. Shin, "Joint optimal sensor selection and scheduling in dynamic spectrum access networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 8, pp. 1532–1545, Aug. 2013.
[19] Z. He and S. Mao, "Adaptive multiple description coding and transmission of uncompressed video over 60 GHz networks," *ACM Mobile Comput. Commun. Rev.*, vol. 18, no. 1, pp. 14–24, Jan. 2014.
[20] Y. Chen, Y. Wu, B. Wang, and K. J. R. Liu, "Spectrum auction games for multimedia streaming over cognitive radio networks," *IEEE Trans. Commun.*, vol. 58, no. 8, pp. 2381–2390, Aug. 2010.
[21] S. Ali and F. Yu, "Cross-layer QoS provisioning for multimedia transmissions in cognitive radio networks," *in Proc. IEEE Wireless Commun. Netw. Conf.*, Budapest, Hungary, Apr. 2009, pp. 1–5.
[22] G. B. Dantzig and P. Wolfe, "The decomposition principle for linear programs," *Oper. Res.*, vol. 8, no. 1, pp. 101–111, Feb. 1960.
[23] J. J. Lyu, H. Luh, and M. C. Lee, "Solving linear programming problems on the parallel virtual machine environment," *Amer. J. Appl. Sci.*, vol. 1, no. 2, pp. 90–94, 2004.
[24] R. Garfinkel and G. L. Nemhauser, *Integer Programming*. Hoboken, NJ, USA: Wiley, 1972.
[25] I. Griva, S. G. Nash, and A. Sofer, Linear and Nonlinear Optimization, 2nd ed. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2009.
[26] T. Jiang, H. Wang, and A. Vasilakos, "QoE-driven channel allocation schemes for multimedia transmission of priority-based secondary users over cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 7, pp. 1215–1224, Aug. 2012.
[27] A. Khan, L. Sun, and E. Ifeachor, "QoE prediction model and its application in video quality adaptation over UMTS networks," *IEEE Trans. Multimedia*, vol. 14, no. 2, pp. 431–442, Apr. 2012.
[28] Xiph.org foundation, *Video Test Media* [derf's collection] [Online]. Available: https://media.xiph.org/video/derf/
[29] D. Hu and S. Mao, "On medium grain scalable video streaming over cognitive radio femtocell networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 641–651, Apr. 2012.

**Zhifeng He** (S'15) received the B.S. degree in electronics information science and technology from Shandong University of Technology, Zibo, China, and the M.S. degree in micro electronics and solid state electronics from Beijing University of Posts and Telecommunications, Beijing, China, in 2009 and 2012, respectively. Since 2012, he has been pursuing the Ph.D. degree in electrical and computer engineering at Auburn University, Auburn, AL, USA. His research interests include cognitive radio, mmWave networks, multimedia streaming, and optimization.

**Shiwen Mao** (S'99–M'04–SM'09) received the Ph.D. degree in electrical and computer engineering from the Polytechnic University (now New York University Polytechnic School of Engineering), Brooklyn, NY, USA, in 2004. He is the Samuel Ginn Endowed Professor and Director of the Wireless Engineering Research and Education Center (WEREC), Auburn University, Auburn, AL, USA. His research interests include wireless networks and multimedia communications. He is on the Editorial Board of IEEE TRANSACTIONS ON MULTIMEDIA, IEEE INTERNET OF THINGS JOURNAL, IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, IEEE MULTIMEDIA, among others. He serves as Steering Committee Member for the IEEE ICME and AdhocNets, area TPC Chair of IEEE INFOCOM 2016, Technical Program Vice Chair for Information Systems of IEEE INFOCOM 2015, and symposium/track Co-Chair for many conferences, including the IEEE ICC, the IEEE GLOBECOM, and the ICCCN. He is a Distinguished Lecturer of the IEEE Vehicular Technology Society, and the Vice Chair—Letters and Member Communications of IEEE Communications Society Multimedia Communications Technical Committee. He was the recipient of the 2013 IEEE ComSoc MMTC Outstanding Leadership Award and the NSF CAREER Award in 2010, and also a co-recipient of the IEEE WCNC 2015 Best Paper Award, the IEEE ICC 2013 Best Paper Award, and the 2004 IEEE Communications Society Leonard G. Abraham Prize in the Field of Communications Systems.

**Sastry Kompella** (S'04–M'07–SM'12) received the B.E. degree in electronics and communication engineering from Andhra University, Vishakhapatnam, India, the M.S. degree in electrical engineering from Texas Tech University, Lubbock, TX, USA, and the Ph.D. degree in electrical and computer engineering from the Virginia Polytechnic Institute and State University, Blacksburg, VA, USA, in 1996, 1998, and 2006, respectively. Currently, he is the Head of the wireless network research section under the Information Technology Division, U.S. Naval Research Laboratory, Washington, DC, USA. His research interests include wireless networks, from mobile *ad hoc* to underwater acoustic networks, with specific focus toward the development of cognitive and cooperative network optimization techniques for efficient wireless link scheduling.