Joint Optimal Cooperative Sensing and Resource Allocation in Multichannel Cognitive Radio Networks

Article *in* IEEE Transactions on Vehicular Technology · March 2011

DOI: 10.1109/TVT.2010.2096545 · Source: IEEE Xplore

Rongfei Fan
Beijing Institute of Technology

Hai Jiang
Arkansas State University - Jonesboro

in which I_m can be expressed as

$$\begin{aligned} I_m &= 1 + \sum_{k=1}^{M-1} \sum_{\substack{A_k \subseteq \{1,2,\ldots,m-1,m+1,\ldots,M\} \\ |A_k|=k}} (-1)^k \prod_{j \in A_k} e^{-y\lambda_{S_j D}} \\ &= 1 + \sum_{k=1}^{M-1} \sum_{\substack{A_k \subseteq \{1,2,\ldots,m-1,m+1,\ldots,M\} \\ |A_k|=k}} (-1)^k e^{-y\sum_{j \in A_k} \lambda_{S_j D}} \end{aligned} \tag{26}$$

where we have used the multinomial expansion identity [16, eq. (33)]. By substituting (26) into (25) and performing the required integral, a closed-form expression is attained for $\Pr(m^* = m)$, as shown in (17).

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and the anonymous reviewers for their insightful suggestions, which greatly improved the quality and presentation of this paper. H. Ding would like to thank Dr. H. Guo for valuable discussions.

REFERENCES

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

[2] A. Bletsas, H. Shin, and M. Z. Win, "Cooperative communications with outage-optimal opportunistic relaying," *IEEE Trans. Wirel. Commun.*, vol. 6, no. 9, pp. 3450–3460, Sep. 2007.

[3] E. G. Larsson, "On the combination of spatial diversity and multiuser diversity," *IEEE Commun. Lett.*, vol. 8, no. 8, pp. 517–519, Aug. 2004.

[4] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.

[5] Q. Zhou and H. Dai, "Asymptotic analysis on the interaction between spatial diversity and multiuser diversity in wireless networks," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4271–4283, Aug. 2007.

[6] H. J. Joung and C. Mun, "Capacity of multiuser diversity with cooperative relaying in wireless networks," *IEEE Commun. Lett.*, vol. 12, no. 10, pp. 752–754, Oct. 2008.

[7] Y. U. Jang, W. Y. Shin, and Y. H. Lee, "Multiuser scheduling based on reduced feedback information in cooperative communications," in *Proc. IEEE VTC-Spring*, Barcelona, Spain, Apr. 2009, pp. 1341–1345.

[8] X. Zhang, W. Wang, and X. Ji, "Multiuser diversity in multiuser two-hop cooperative relay wireless networks: System model and performance analysis," *IEEE Trans. Veh. Technol.*, vol. 58, no. 2, pp. 1031–1036, Feb. 2009.

[9] S. Chen, W. Wang, and X. Zhang, "Performance analysis of multiuser diversity in cooperative multi-relay networks under Rayleigh-fading channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3415–3419, Jul. 2009.

[10] L. Sun, T. Zhang, L. Lu, and H. Niu, "On the combination of cooperative diversity and multiuser diversity in multi-source multi-relay wireless networks," *IEEE Signal Process. Lett.*, vol. 17, no. 6, pp. 535–538, Jun. 2010.

[11] Y. Zhao, R. Adve, and T. J. Lim, "Improving amplify-and-forward relay networks: Optimal power allocation versus selection," *IEEE Trans. Wirel. Commun.*, vol. 6, no. 8, pp. 3114–3123, Aug. 2007.

[12] Y. Zhao, R. Adve, and T. J. Lim, "Symbol error rate of selection amplify-and-forward relay systems," *IEEE Commun. Lett.*, vol. 10, no. 11, pp. 757–759, Nov. 2006.

[13] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 4th ed. New York: McGraw-Hill, 2002.

[14] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products,* 7th ed. San Diego, CA: Academic, 2007.

[15] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1972.

[16] A. Bletsas, A. G. Dimitriou, and J. N. Sahalos, "Interference-limited opportunistic relaying with reactive sensing," *IEEE Trans. Wireless Commun.*, vol. 9, no. 1, pp. 14–20, Jan. 2010.

Joint Optimal Cooperative Sensing and Resource Allocation in Multichannel Cognitive Radio Networks

Rongfei Fan, *Student Member, IEEE*, Hai Jiang, *Member, IEEE*, Qiang Guo, and Zhou Zhang

***Abstract*—In this paper, the problem of joint multichannel cooperative sensing and resource allocation is investigated. A cognitive radio network with multiple potential channels and multiple secondary users is considered. Each secondary user carries out wideband spectrum sensing to get a test statistic for each channel and transmits the test statistic to a coordinator. After collecting all the test statistics from secondary users, the coordinator makes the estimation as to whether primary users are idle or not in the channels. When a channel is estimated to be free, secondary users can get access to the channel with assigned bandwidth and power. An optimization problem is formulated, which maximizes the weighted sum of secondary users' throughputs while guaranteeing a certain level of protection for the activities of primary users. Although the problem is nonconvex, it is shown that the problem can be solved by bilevel optimization and monotonic programming. This paper is also extended to cases with consideration of proportional and max–min fairness.**

***Index Terms*—Cognitive radio, resource allocation, spectrum sensing.**

I. INTRODUCTION

As a viable solution for future spectrum-scarcity problems, the *cognitive radio* technique employs the idea of *opportunistic spectrum access* in which unlicensed (secondary) users can utilize the spectrum when the licensed (primary) users are detected (usually by spectrum sensing) not using the spectrum [1]–[10]. A slotted structure is broadly adopted in the literature [3]–[8] for cognitive radio networks. Each slot is divided into two parts, namely, spectrum sensing and data transmission. Secondary users sense the spectrum in the spectrum-sensing part to check whether there are primary users using the spectrum and transmit in the data-transmission part if the spectrum is idle. In a cognitive radio network, it is likely that multiple channels are available. Then, it is frequently assumed in the literature that a secondary user can sense only one channel at a time [3]–[6], [8], [11]–[13]. To improve sensing accuracy in a fading environment, cooperative spectrum sensing can be used [14]. In cooperative spectrum sensing, multiple secondary users independently sense the spectrum and deliver their sensing results to a coordinator, which makes the

Manuscript received March 24, 2010; revised September 20, 2010 and November 9, 2010; accepted November 26, 2010. Date of publication December 3, 2010; date of current version February 18, 2011. This work was supported in part by the Natural Science and Engineering Research Council (NSERC) of Canada and in part by the Alberta Innovates—Technology Futures, Alberta, Canada. The review of this paper was coordinated by Prof. B. Hamdaoui.

The authors are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada (e-mail: rongfei@ualberta.ca; hai.jiang@ece.ualberta.ca; qguo@ualberta.ca; zzhang6@ualberta.ca).

Digital Object Identifier 10.1109/TVT.2010.2096545

final decision on whether the spectrum is available. Two parameters are generally used to measure the performance of spectrum sensing, namely, detection probability (the probability of detecting the activity of a primary user when the primary user is active) and false-alarm probability (the probability of mistakenly claiming that a primary user is active when the primary user is actually idle). Generally speaking, the duration of the spectrum-sensing part in a slot determines the accuracy of spectrum sensing. A longer spectrum-sensing duration leads to a higher detection probability and a lower false-alarm probability. However, it also means less time in the slot for the data-transmission part. The optimal tradeoff in the configuration of the spectrum-sensing duration is investigated in [8] and [15] for the case of a single channel and in [11] for the case of multiple channels.

In those research efforts dealing with the configuration of the spectrum-sensing duration [8], [11], [15], it is assumed that each secondary user has a fixed transmission power and rate. In other words, the gain of dynamic resource allocation is not exploited. In addition, it is assumed that each user can sense or access only one channel at a time. Recently, the wideband sampling technique has received some initial investigation [16], which makes wideband spectrum sensing and access possible. All these have motivated us to investigate the joint optimization of spectrum-sensing duration configuration and resource allocation with wideband spectrum sensing and access.

The rest of this paper is organized as follows: In Section II, the system model is given. A weighted throughput optimization problem is formulated in Section III. Although it is not a convex problem, a combination of bilevel optimization and monotonic programming methods is used to solve the problem. Further extensions with consideration of proportional and max–min fairness are presented in Section IV. Numerical results are presented in Section V, followed by concluding remarks in Section VI.

II. System Model

A cognitive radio network is considered with N frequency bands (termed *channels* in the sequel) and M secondary communication pairs (with the mth pair including secondary transmitter m and secondary receiver m). Each channel has a bandwidth of W and is licensed to one primary user.

The system is assumed to be perfectly synchronized, and time is divided into slots, each with a fixed length T. In each slot, the primary user in a channel is assumed to be either active or idle for the whole slot. The channels among primary and secondary users are assumed to keep unchanged within each time slot. In other words, a block-fading model is assumed, and the channel coherent time is assumed to be longer than the slot duration T. This assumption is reasonable when all the primary and secondary users in the system are static or in low-speed mobility.

Each slot includes three phases: 1) a sensing phase; 2) a reporting phase; and 3) a transmission phase. The duration of the sensing phase is denoted τ, which is a parameter to be optimized. In the sensing phase, each secondary user senses the signal spanning the N channels, with a sampling rate μ. In the reporting phase, the secondary users report to a coordinator their test statistics in the sensing phase and their channel gains. Based on these, the coordinator determines how the secondary users can get access to the channels and broadcasts the resource-allocation decision to secondary users. During the reporting phase, the information exchange between secondary users and the coordinator is on a common control channel, for example, on the industrial, scientific, and medical (ISM) radio bands. In addition, to keep the duration of the reporting phase at an acceptable level, the number of secondary users that are involved should be limited [17]. It can be seen that the reporting phase adds a fixed time overhead. Without loss of generality, the duration of the reporting phase is assumed to be zero in the sequel. The transmission phase is used for data transmission of secondary users.

A. Spectrum Sensing

By the wideband sampling technique [16], for secondary transmitter m, its received primary signal at channel n in the frequency domain, i.e., $R_{n,m}$, can be given in the following binary test hypothesis:

$$\begin{aligned}\mathcal{H}_n^0 &: \ R_{n,m} = V_{n,m}\\ \mathcal{H}_n^1 &: \ R_{n,m} = H_{n,m}S_n + V_{n,m}\end{aligned} \tag{1}$$

where $\mathcal{H}_n^0$ and $\mathcal{H}_n^1$ mean the primary user on channel n is idle and busy, respectively; and $V_{n,m}$, $H_{n,m}$, and S_n are N-point fast Fourier transformations (FFTs) of the additive complex white Gaussian noise (with mean being zero and variance being σ^2), the discrete-time channel impulse response, and the primary signal, respectively. For presentation simplicity, when the primary signal is present, the power spectrum density of the primary signal on each channel at the transmitter side is normalized to be 1, i.e., $\mathbb{E}\{|S_n|^2\} = 1$, where $\mathbb{E}\{\cdot\}$ denotes expectation.

Within a duration of τ in the sensing phase, each user collects $\mu\tau$ samples of the received signal for all N channels. By doing FFT on every sampled N points, we have a number $\mu\tau/N$ of values for $R_{n,m}$, which are denoted $R_{n,m}(1), R_{n,m}(2), \ldots, R_{n,m}(\mu\tau/N)$. Employing energy detection, secondary user m has a *test statistic* for channel n, which is given as

$$T_{n,m} \triangleq \frac{N}{\mu\tau}\sum_{k=1}^{\mu\tau/N} |R_{n,m}(k)|^2 . \tag{2}$$

The coordinator collects test statistics $T_{n,m}$'s from all secondary users and makes estimation whether the primary signal in channel n exists. If the detection threshold of channel n is ε_n, then the detection and false-alarm probabilities for channel n are given, respectively, by following the way in [16] as

$$P_n^d(\tau,\varepsilon_n) = Q\left(\frac{\left(\varepsilon_n - (\sigma^2 M + \|\boldsymbol{H}_n\|_1)\right)\sqrt{\mu\tau}}{\sigma\sqrt{2\|\boldsymbol{\Sigma}_n\|_1 N}}\right) \tag{3}$$

$$P_n^f(\tau,\varepsilon_n) = Q\left(\frac{(\varepsilon_n - \sigma^2 M)\sqrt{\mu\tau}}{\sigma^2\sqrt{2MN}}\right) \tag{4}$$

where $\boldsymbol{H}_n = (|H_{n,1}|^2, |H_{n,2}|^2, \ldots, |H_{n,M}|^2)^T$ (here, superscript T means transpose operation), $\boldsymbol{\Sigma}_n = \sigma^2\boldsymbol{I} + 2\mathrm{Diag}(\boldsymbol{H}_n)$ (here, $\boldsymbol{I}$ means an identity matrix, and $\mathrm{Diag}(\boldsymbol{H}_n)$ is a diagonal matrix formed by elements in $\boldsymbol{H}_n$), $Q(\cdot)$ is the Q-function, which is defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{y^2}{2}}\,dy \tag{5}$$

and $\|\cdot\|_1$ of a matrix represents its entrywise 1-norm, which is the summation of absolute values of all the elements. Based on (3) and (4), P_n^f can be expressed in terms of P_n^d as

$$P_n^f\left(\tau, P_n^d\right) = Q\left(\frac{Q^{-1}\left(P_n^d\right)\sigma\sqrt{2\|\boldsymbol{\Sigma}_n\|_1 N} + \|\boldsymbol{H}_n\|_1\sqrt{\mu\tau}}{\sigma^2\sqrt{2MN}}\right). \tag{6}$$

In a real system, the false-alarm probability should be no larger than 0.5. For a given value of P_n^d, constraint $P_n^f \le 0.5$ is equivalent to the following constraint:

$$Q^{-1}\left(P_n^d\right)\sigma\sqrt{2\|\boldsymbol{\Sigma}_n\|_1 N}+\|\boldsymbol{H}_n\|_1\sqrt{\mu\tau}\geq 0 \tag{7}$$

which is from (6).

B. Resource Allocation

Recall that the coherence time of the channels (from primary to secondary users and among secondary users) is assumed to be relatively long. Within the duration of coherence time, the resource-allocation strategy does not change. For the mth secondary user pair in channel n, denote the channel gain from the primary user to the secondary receiver and from the secondary transmitter to the secondary receiver by $|g_{n,m}^{\mathrm{ps}}|^2$ and $|g_{n,m}^{\mathrm{ss}}|^2$, respectively.

In resource allocation, each secondary user is assigned a portion of each channel at a certain transmission power. Specifically, secondary user m is assigned a portion $x_{n,m}(0\leq x_{n,m}\leq 1)$ of channel n[1] at a transmission power $p_{n,m}$. If channel n is estimated by the coordinator to be free, then secondary user m can transmit in the assigned portion of channel n with the assigned power level. It is assumed that the channel gain in channel n between the mth secondary transceiver pair, i.e., $|g_{n,m}^{\mathrm{ss}}|^2$, is a constant within the bandwidth of channel n. Therefore, it does not make any difference which portion in channel n is assigned to secondary user m.

When channel n is estimated (by the coordinator) to be free, one of two scenarios happens.

1) Channel n is indeed free (i.e., the estimation is correct), with probability $\Pr(\mathcal{H}_n^0)(1-P_n^f(\tau,\varepsilon_n))$, where $\Pr(\mathcal{H}_n^0)$ is the free probability of channel n. The achievable transmission rate of secondary user m is

$$r_{n,m}^0=x_{n,m}\cdot\log\left(1+\frac{\left|g_{n,m}^{\mathrm{ss}}\right|^2 p_{n,m}}{Wx_{n,m}\sigma^2}\right).$$

2) Channel n is actually busy (i.e., the estimation is wrong), with probability $\Pr(\mathcal{H}_n^1)(1-P_n^d(\tau,\varepsilon_n))$, where $\Pr(\mathcal{H}_n^1)=1-\Pr(\mathcal{H}_n^0)$ is the busy probability of channel n. In this scenario, the primary signal serves as interference to secondary transmission. The achievable transmission rate of secondary user m is

$$r_{n,m}^1=x_{n,m}\cdot\log\left(1+\frac{\left|g_{n,m}^{\mathrm{ss}}\right|^2 p_{n,m}}{x_{n,m}W\left(|g_{n,m}^{\mathrm{ps}}|^2+\sigma^2\right)}\right).$$

Then, the average throughput of secondary user m is given as

$$R_m=\left(1-\frac{\tau}{T}\right)\cdot\sum_{n=1}^{N}\left[\Pr\left(\mathcal{H}_n^0\right)\left(1-P_n^f(\tau,\varepsilon_n)\right)\cdot r_{n,m}^0 + \Pr\left(\mathcal{H}_n^1\right)\left(1-P_n^d(\tau,\varepsilon_n)\right)\cdot r_{n,m}^1\right]. \tag{8}$$

Suppose the average transmission power of secondary transmitter m is bounded by P_m^{avg} and the maximal instantaneous transmission power of secondary transmitter m is bounded by P_m^{peak}. Then, for secondary transmitter m, we have the following constraints for resource allocation:

$$\sum_{n=1}^{N}p_{n,m}\leq P_m^{\mathrm{peak}},\quad \sum_{n=1}^{N}\Pr\left(\mathcal{H}_n^0\right)p_{n,m}\leq P_m^{\mathrm{avg}}. \tag{9}$$

[1]This can be achieved via the orthogonal frequency-division multiple-access technique by partitioning each channel into a number of subcarriers.

III. Joint Optimization of Sensing-Time Setting and Resource Allocation

In this section, a joint sensing-duration configuration and resource-allocation optimization problem is formulated and addressed, which maximizes the weighted sum of average throughputs of all the secondary users, given as $R_w=\sum_{m=1}^{M}\alpha_m\cdot R_m$, where $\alpha_m\geq 0$ is the weight assigned to secondary user m.

To protect the activities of primary users, the detection probability in each channel should be larger than a threshold that is denoted P_{th} (which should be much larger than 0.5). Together with the constraints discussed in the preceding section, an optimization problem can be formulated.

Problem P1:

$$\underset{\tau,\{\varepsilon_n\},\{p_{n,m}\},\{x_{n,m}\}}{\text{maximize}}$$

$$R_w\left(\tau,\{\varepsilon_n\},\{p_{n,m}\},\{x_{n,m}\}\right)$$
$$=\left(1-\frac{\tau}{T}\right)\cdot\sum_{n=1}^{N}\left[\Pr\left(\mathcal{H}_n^0\right)\cdot\left(1-P_n^f(\tau,\varepsilon_n)\right)\cdot\sum_{m=1}^{M}\alpha_m r_{n,m}^0+\Pr\left(\mathcal{H}_n^1\right)\cdot\left(1-P_n^d(\tau,\varepsilon_n)\right)\cdot\sum_{m=1}^{M}\alpha_m r_{n,m}^1\right]$$

subject to

$$0\leq\tau\leq T \tag{10a}$$
$$P_n^d(\tau,\varepsilon_n)\geq P_{\mathrm{th}},\qquad n\in\mathcal{N} \tag{10b}$$
$$Q^{-1}\left(P_n^d(\tau,\varepsilon_n)\right)\sigma\sqrt{2\|\boldsymbol{\Sigma}_n\|_1 N}+\|\boldsymbol{H}_n\|_1\sqrt{\mu\tau}\geq 0,\qquad n\in\mathcal{N} \tag{10c}$$
$$\sum_{m=1}^{M}x_{n,m}=1,\qquad n\in\mathcal{N} \tag{10d}$$
$$\sum_{n=1}^{N}p_{n,m}\leq P_m^{\mathrm{peak}},\qquad m\in\mathcal{M} \tag{10e}$$
$$\sum_{n=1}^{N}\Pr\left(\mathcal{H}_n^0\right)p_{n,m}\leq P_m^{\mathrm{avg}},\qquad m\in\mathcal{M} \tag{10f}$$
$$0\leq x_{n,m}\leq 1,\qquad n\in\mathcal{N},\ m\in\mathcal{M} \tag{10g}$$
$$0\leq p_{n,m}\leq P_m^{\mathrm{peak}},\qquad n\in\mathcal{N},\ m\in\mathcal{M} \tag{10h}$$

where $\mathcal{N}=\{1,2,\ldots,N\}$, and $\mathcal{M}=\{1,2,\ldots,M\}$.

Problem P1 is nonconvex. It can be proved that Problem P1 achieves the optimal solution only when $P_n^d(\tau,\varepsilon_n)=P_{\mathrm{th}}$. The proof is omitted, as it is similar to that in [11, Lemma 2].

Then, in Problem P1, the term $P_n^d(\tau,\varepsilon_n)$ can be replaced with P_{th}, and $P_n^f(\tau,\varepsilon_n)$ can be expressed in terms of τ and P_{th} according to (6). Note that, by replacing $P_n^d(\tau,\varepsilon_n)$ with P_{th}, constraint (10c) becomes

$$Q^{-1}(P_{\mathrm{th}})\sigma\sqrt{2\|\boldsymbol{\Sigma}_n\|_1 N}+\|\boldsymbol{H}_n\|_1\sqrt{\mu\tau}\geq 0$$

which is equivalent to

$$\tau \geq \tau_{\min}^{n} \triangleq \frac{2\left(Q^{-1}(P_{\mathrm{th}})\right)^{2}\sigma^{2}\|\boldsymbol{\Sigma}_{n}\|_{1}N}{\mu\|\boldsymbol{H}_{n}\|_{1}^{2}}.$$

Denote $\tau_{\min} = \max(\tau_{\min}^{1}, \tau_{\min}^{2}, \ldots, \tau_{\min}^{N})$. Then, Problem P1 is equivalent to the following problem:

Problem P2:

$$\underset{\tau,\{p_{n,m}\},\{x_{n,m}\}}{\text{maximize}}$$

$$R_{w}\left(\tau,\{p_{n,m}\},\{x_{n,m}\}\right) = \left(1-\frac{\tau}{T}\right) \cdot \sum_{n=1}^{N}\left(\Pr\left(\mathcal{H}_{n}^{0}\right)\cdot\left(1-P_{n}^{f}(\tau,P_{\mathrm{th}})\right)\cdot\sum_{m=1}^{M}\alpha_{m}r_{n,m}^{0} + \Pr\left(\mathcal{H}_{n}^{1}\right)\cdot(1-P_{\mathrm{th}})\cdot\sum_{m=1}^{M}\alpha_{m}r_{n,m}^{1}\right)$$

subject to

$$\tau_{\min} \leq \tau \leq T \tag{11a}$$

$$\text{Constraints (10d) – (10h).} \tag{11b}$$

Problem P2 is still a nonconvex problem. To solve Problem P2, we use *bilevel optimization*, in which the lower level problem is to optimize $\{p_{n,m}\}$ and $\{x_{n,m}\}$ with a fixed τ, whereas the upper level problem is to optimize τ. Specifically, the lower level problem is

Problem P3:

$$U(\tau) = \max_{\{p_{n,m}\},\{x_{n,m}\}} \sum_{n=1}^{N}\left[\Pr\left(\mathcal{H}_{n}^{0}\right)\cdot\left(1-P_{n}^{f}(\tau,P_{\mathrm{th}})\right) \cdot\sum_{m=1}^{M}\alpha_{m}r_{n,m}^{0} + \Pr\left(\mathcal{H}_{n}^{1}\right)\cdot(1-P_{\mathrm{th}})\cdot\sum_{m=1}^{M}\alpha_{m}r_{n,m}^{1}\right]$$

subject to constraints (10d)–(10h).

Denote $V(\tau) = (\tau/T)U(\tau)$. Then, the upper level problem is

Problem P4:

$$\underset{\tau}{\text{maximize}} \quad U(\tau) - V(\tau)$$

$$\text{subject to} \quad \tau_{\min} \leq \tau \leq T.$$

The lower level problem, i.e., Problem P3, is convex [18]. Thus, the optimal solution of $\{p_{n,m}\}$ and $\{x_{n,m}\}$ for a given τ can be obtained.

For the upper level problem, i.e., Problem P4, we have the following lemma.

Lemma 1: Functions $U(\tau)$ and $V(\tau)$ are monotonically increasing functions with respect to τ within the interval $[\tau_{\min}, T]$.

Proof: We first prove that $U(\tau)$ is a monotonically increasing function with respect to τ.

Suppose $\tau^{\dagger} < \tau^{\ddagger}$ and the optimal solutions of Problem P3 with $\tau = \tau^{\dagger}$ and $\tau = \tau^{\ddagger}$ are $\{p_{n,m}^{\dagger}\}$, $\{x_{n,m}^{\dagger}\}$, and $\{p_{n,m}^{\ddagger}\}$, $\{x_{n,m}^{\ddagger}\}$, respectively. Since $(1-P_{n}^{f}(\tau,P_{\mathrm{th}}))$ is a monotonically increasing function with respect to τ within the interval $[\tau_{\min}, T]$ according to (6), we have

$$\begin{aligned}
U(\tau^{\dagger}) &= \sum_{n=1}^{N}\left[\Pr\left(\mathcal{H}_{n}^{0}\right)\cdot\left(1-P_{n}^{f}(\tau^{\dagger},P_{\mathrm{th}})\right) \cdot\sum_{m=1}^{M}\alpha_{m}x_{n,m}^{\dagger}\log\left(1+\frac{\left|g_{n,m}^{\mathrm{ss}}\right|^{2}p_{n,m}^{\dagger}}{Wx_{n,m}^{\dagger}\sigma^{2}}\right) + \Pr\left(\mathcal{H}_{n}^{1}\right)\cdot(1-P_{\mathrm{th}}) \cdot\sum_{m=1}^{M}\alpha_{m}x_{n,m}^{\dagger}\log\left(1+\frac{\left|g_{n,m}^{\mathrm{ss}}\right|^{2}p_{n,m}^{\dagger}}{Wx_{n,m}^{\dagger}\left(|g_{n,m}^{\mathrm{ps}}|^{2}+\sigma^{2}\right)}\right)\right]\\
&< \sum_{n=1}^{N}\left[\Pr\left(\mathcal{H}_{n}^{0}\right)\cdot\left(1-P_{n}^{f}(\tau^{\ddagger},P_{\mathrm{th}})\right) \cdot\sum_{m=1}^{M}\alpha_{m}x_{n,m}^{\dagger}\log\left(1+\frac{\left|g_{n,m}^{\mathrm{ss}}\right|^{2}p_{n,m}^{\dagger}}{Wx_{n,m}^{\dagger}\sigma^{2}}\right) + \Pr\left(\mathcal{H}_{n}^{1}\right)\cdot(1-P_{\mathrm{th}}) \cdot\sum_{m=1}^{M}\alpha_{m}x_{n,m}^{\dagger}\log\left(1+\frac{\left|g_{n,m}^{\mathrm{ss}}\right|^{2}p_{n,m}^{\dagger}}{Wx_{n,m}^{\dagger}\left(|g_{n,m}^{\mathrm{ps}}|^{2}+\sigma^{2}\right)}\right)\right]\\
&\leq \sum_{n=1}^{N}\left[\Pr\left(\mathcal{H}_{n}^{0}\right)\cdot\left(1-P_{n}^{f}(\tau^{\ddagger},P_{\mathrm{th}})\right) \cdot\sum_{m=1}^{M}\alpha_{m}x_{n,m}^{\ddagger}\log\left(1+\frac{\left|g_{n,m}^{\mathrm{ss}}\right|^{2}p_{n,m}^{\ddagger}}{Wx_{n,m}^{\ddagger}\sigma^{2}}\right) + \Pr\left(\mathcal{H}_{n}^{1}\right)\cdot(1-P_{\mathrm{th}}) \cdot\sum_{m=1}^{M}\alpha_{m}x_{n,m}^{\ddagger}\log\left(1+\frac{\left|g_{n,m}^{\mathrm{ss}}\right|^{2}p_{n,m}^{\ddagger}}{Wx_{n,m}^{\ddagger}\left(|g_{n,m}^{\mathrm{ps}}|^{2}+\sigma^{2}\right)}\right)\right]\\
&= U(\tau^{\ddagger})
\end{aligned}$$

where the last inequality comes from the fact that the optimal solution of Problem P3 when $\tau = \tau^{\ddagger}$ is $\{p_{n,m}^{\ddagger}\}$ and $\{x_{n,m}^{\ddagger}\}$. Therefore, $U(\tau)$ is monotonically increasing. Furthermore, $V(\tau)$ is also monotonically increasing since it is the product of two positive monotonically increasing functions, i.e., (τ/T) and $U(\tau)$. ■

Denote $\tau' = \tau - \tau_{\min}$. By introducing a new variable ω, Problem P4 is equivalent to

Problem P5:

$$\begin{aligned}
\underset{\tau',\omega}{\text{maximize}} \quad & U(\tau'+\tau_{\min}) + \omega\\
\text{subject to} \quad & \omega + V(\tau'+\tau_{\min}) \leq V(T)\\
& 0 \leq \tau' \leq T - \tau_{\min}\\
& \omega \geq 0.
\end{aligned}$$

The reason for the equivalence of Problems P4 and P5 is as follows: In Problem P5, the objective function $(U(\tau' + \tau_{\min}) + \omega)$ is a monotonically increasing function. The function $(\omega + V(\tau' + \tau_{\min}))$ is also monotonically increasing. Therefore, the maximal objective function happens only when $\omega + V(\tau' + \tau_{\min}) = V(T)$. This means that Problem P5 is to maximize

$$\begin{aligned} U(\tau' + \tau_{\min}) + \omega &= U(\tau' + \tau_{\min}) - V(\tau' + \tau_{\min}) + V(T) \\ &= U(\tau) - V(\tau) + V(T) \end{aligned}$$

which is equivalent to maximizing $(U(\tau) - V(\tau))$ since $V(T)$ is a constant. Thus, Problems P4 and P5 are equivalent.

Problem P5 is in the form of a *monotonic optimization problem* [20] because it has the following features: 1) The object function, i.e., $(U(\tau' + \tau_{\min}) + \omega)$, is a monotonically increasing function; 2) all constraint functions are monotonic functions; and 3) both τ' and ω lie in the region of $[0, +\infty)$. A monotonic optimization problem can be solved by a monotonic programming method, named *polyblock algorithm* [20] with parameter ϵ, which is ϵ-optimal (i.e., the difference of the achieved utility in the polyblock algorithm from the global optimal utility is bounded by ϵ). To be specific, for Problem P5, a polyblock algorithm is given as follows. See [20] for a detailed discussion of the algorithm.

Polyblock Algorithm

1: Define a point set $\mathcal{S} = \{s_1, s_2, \ldots, s_{|\mathcal{S}|}\}$, in which each point has two elements. Initialize set $\mathcal{S}$ by one point $s_1 = (T - \tau_{\min}, V(T) - V(\tau_{\min}))$. In other words, the two elements in point s_1 are $s_1(1) = T - \tau_{\min}$ and $s_1(2) = V(T) - V(\tau_{\min})$.
2: **while** Set $\mathcal{S} \neq \emptyset$ **do**
3: **for** $i = 1, 2, \ldots, |\mathcal{S}|$ **do**
4: Calculate λ_i that satisfies $\lambda_i s_i(2) + V(\lambda_i s_i(1) + \tau_{\min}) = V(T)$ by a bisection search, and set $\pi_i = \lambda_i s_i$.
5: Find $i^* = \arg\max_{1 \le i \le |\mathcal{S}|} [U(\pi_i(1) + \tau_{\min}) + \pi_i(2)]$
6: For any point (e.g., point s) in set $\mathcal{S}$, if $[U(s(1) + \tau_{\min}) + s(2)] \le [U(\pi_{i^*}(1) + \tau_{\min}) + \pi_{i^*}(2)] + \epsilon$, then remove the point from $\mathcal{S}$.
7: **if** Set $\mathcal{S} \neq \emptyset$ **then**
8: Find $j^* = \arg\max_{1 \le j \le |\mathcal{S}|} [U(s_j(1) + \tau_{\min}) + s_j(2)]$
9: Calculate λ_{j^*} that satisfies $\lambda_{j^*} s_{j^*}(2) + V(\lambda_{j^*} s_{j^*}(1) + \tau_{\min}) = V(T)$ by a bisection search, and set $\pi_{j^*} = \lambda_{j^*} s_{j^*}$.
10: Generate two points $s^\dagger = s_{j^*} + (\pi_{j^*} - s_{j^*}) \circ (1, 0)$ and $s^\ddagger = s_{j^*} + (\pi_{j^*} - s_{j^*}) \circ (0, 1)$, where $\circ$ means Hadamard product.
11: Add $s^\dagger$ and $s^\ddagger$ into $\mathcal{S}$, delete s_{j^*} from $\mathcal{S}$.
12: Output the last π_{i^*} before $\mathcal{S}$ becomes an empty set.

By employing the polyblock algorithm, the first and second elements of π_{i^*} are the optimal τ' and ω, respectively, for Problem P5. Then, the optimal τ in Problem P4 is obtained by $\tau = \pi_{i^*}(1) + \tau_{\min}$.

IV. Fairness Consideration

In the preceding section, the weighted throughput maximization problem is solved with the aid of bilevel optimization and monotonic programming. In this section, we focus on the cases when fairness among users is taken into account. Specifically, proportional and max–min fairness are discussed in the following two subsections, respectively.

A. Proportional Fairness Optimization Problem

In this case, the utility function of secondary user m can be written as $\log R_m$. Therefore, the optimization problem is as follows:

Problem P6:

$$\begin{aligned} &\underset{\tau, \{\varepsilon_n\}, \{p_{n,m}\}, \{x_{n,m}\}}{\text{maximize}} \; R_{\text{pf}}\left(\tau, \{\varepsilon_n\}, \{p_{n,m}\}, \{x_{n,m}\}\right) \\ &= \sum_{m=1}^{M} \log \Bigg(\left(1 - \frac{\tau}{T}\right) \cdot \sum_{n=1}^{N} \Big[\Pr\left(\mathcal{H}_n^0\right) \left(1 - P_n^f(\tau, \varepsilon_n)\right) \\ &\qquad \cdot r_{n,m}^0 + \Pr\left(\mathcal{H}_n^1\right) \left(1 - P_n^d(\tau, \varepsilon_n)\right) \\ &\qquad \cdot r_{n,m}^1 \Big] \Bigg) \end{aligned}$$

subject to constraints (10a)–(10h).

Similar to the preceding section, Problem P6 is nonconvex, and it achieves the optimal solution only when $P_n^d(\tau, \varepsilon_n) = P_{\text{th}}$. After substituting $P_n^d(\tau, \varepsilon_n)$ with P_{th}, Problem P6 is equivalent to a bilevel problem. The lower level problem is to optimize $\{p_{n,m}\}$ and $\{x_{n,m}\}$ with a fixed τ, as

Problem P7:

$$\begin{aligned} Z(\tau) = &\max_{\{p_{n,m}\}, \{x_{n,m}\}} \\ &\sum_{m=1}^{M} \log \Bigg(\sum_{n=1}^{N} \Big[\Pr\left(\mathcal{H}_n^0\right) \cdot \left(1 - P_n^f(\tau, P_{\text{th}})\right) \\ &\qquad \cdot r_{n,m}^0 + \Pr\left(\mathcal{H}_n^1\right) \cdot (1 - P_{\text{th}}) \cdot r_{n,m}^1 \Big] \Bigg) \end{aligned}$$

subject to constraints (10d)–(10h).

According to [19, p. 84], the objective function of Problem P7 can be shown to be a concave function. Therefore, Problem P7 is convex. In addition, the maximal objective function in Problem P7, i.e., $Z(\tau)$, is a monotonically increasing function with respect to τ. The proof is similar to that in Lemma 1 and is omitted here.

The upper level problem is

Problem P8:

$$\begin{aligned} &\underset{\tau}{\text{maximize}} \quad R_{pf}(\tau) = \log\left(1 - \frac{\tau}{T}\right)^M + Z(\tau) \\ &\text{subject to} \quad \tau_{\min} \le \tau \le T. \end{aligned}$$

For expression

$$\left(1 - \frac{\tau}{T}\right)^M = \sum_{l=0}^{M} \binom{M}{l} \left(-\frac{\tau}{T}\right)^l \tag{12}$$

denote $Y_1(\tau)$ as the summation of all the positive items on the right-hand side and $Y_2(\tau)$ as the summation of all the absolution values of negative items on the right-hand side. Thus, we have

$$\left(1 - \frac{\tau}{T}\right)^M = Y_1(\tau) - Y_2(\tau). \tag{13}$$

It is clear that both $Y_1(\tau)$ and $Y_2(\tau)$ are positive monotonically increasing functions with respect to τ. Then, the object function in Problem P8 can be changed to

$$e^{R_{\text{pf}}(\tau)} = Y_1(\tau) \cdot e^{Z(\tau)} - Y_2(\tau) \cdot e^{Z(\tau)}. \tag{14}$$

In the new objective function, both $Y_1(\tau) e^{Z(\tau)}$ and $Y_2(\tau) e^{Z(\tau)}$ are monotonically increasing functions with respect to τ. Therefore, Problem P8 has a form similar to Problem P4 and can be solved by a polyblock algorithm.

B. Max–Min Fairness Optimization Problem

With max–min fairness, the optimization problem can be formulated as

Problem P9:

$$\underset{\tau,\{\varepsilon_n\},\{p_{n,m}\},\{x_{n,m}\}}{\text{maximize}}$$
$$\min_m \left(\left(1-\frac{\tau}{T}\right)\cdot\sum_{n=1}^{N}\left[\Pr\left(\mathcal{H}_n^0\right)\left(1-P_n^f(\tau,\varepsilon_n)\right)\cdot r_{n,m}^0 + \Pr\left(\mathcal{H}_n^1\right)\left(1-P_n^d(\tau,\varepsilon_n)\right)\cdot r_{n,m}^1\right]\right)$$

subject to constraints (10a)–(10h).

For Problem P9, it can also be concluded that the optimal utility is achieved only when $P_n^d(\tau,\varepsilon_n)=P_{\text{th}}$.

For the objective function in Problem P9, we have (15), shown at the bottom of the page. Setting $P_n^d(\tau,\varepsilon_n)=P_{\text{th}}$ and applying the transformation shown in (15), Problem P9 is equivalent to a bilevel optimization problem. The lower level problem is

Problem P10:

$$F(\tau)=\max_{\{p_{n,m}\},\{x_{n,m}\},\Omega}\ \Omega$$
subject to
$$\sum_{n=1}^{N}\left(\Pr\left(\mathcal{H}_n^0\right)\left(1-P_n^f(\tau,P_{\text{th}})\right)\cdot r_{n,m}^0 + \Pr\left(\mathcal{H}_n^1\right)(1-P_{\text{th}})\cdot r_{n,m}^1\right)\geq\Omega,\quad m\in\mathcal{M}$$
Constraints (10d)–(10h).

Problem P10 is convex, and its maximal objective function $F(\tau)$ is a monotonically increasing function with respect to τ.

The upper level problem is to maximize $(1-\tau/T)\cdot F(\tau)$ subject to τ within the region defined by constraint $\tau_{\min}\leq\tau\leq T$. Note that the objective function is a difference between two positive monotonically increasing functions, i.e., $F(\tau)$ and $(\tau/T)F(\tau)$, which is similar to Problem P4. Therefore, the upper level problem can be solved by a polyblock algorithm.

V. Numerical Results

In this section, numerical results are demonstrated to verify our proposed algorithms. The system is set up as follows: There are four primary channels and four secondary users in a cognitive radio network. Each of the channels spans a bandwidth of 1 MHz. The free probabilities of the four channels are 0.9, 0.8, 0.7, and 0.6, respectively. The weights assigned to secondary users, i.e., α_m's, are all set to 1.

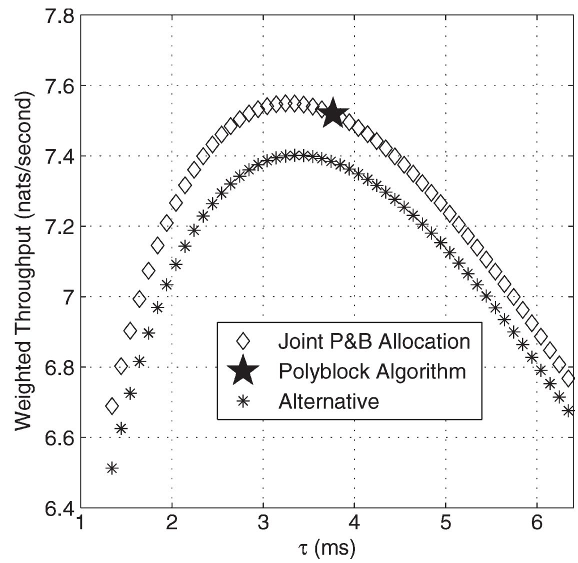

Fig. 1. Weighted sum of throughputs versus τ.

The sampling rate is $\mu=8$ MHz; the slot duration is $T=20$ ms; and the threshold of detection probability is $P_{\text{th}}=0.9$. Of the four secondary users, the average transmission power values P_m^{avg}'s are set to 0.5, 0.45, 0.4, and 0.35, respectively, and the peak transmission power values P_m^{peak}'s are set to 0.8, 0.9, 1, and 1.2, respectively. All links (from primary to secondary users and between secondary users) experience Rayleigh fading. Furthermore, in each channel, the channel signal-to-noise ratio (SNR) value between each secondary transceiver pair is with mean 15 dB, and the channel SNR value from the primary user to a secondary user (either the transmitter or the receiver) is with mean -15 dB. The value of ϵ is set to 0.05.

We first demonstrate the effect of the sensing-time duration configuration in our research. For different particular values of τ in each of the three optimization problems in Sections III and IV, we obtain the optimal utilities under the method of joint power and bandwidth allocation (which is a convex optimization problem). The corresponding curves are marked as "Joint P&B Allocation" in Figs. 1–3. It can be seen that, as τ grows, the optimal utilities first increase and then decrease. The first increase in the optimal utilities is due to the improvement in false-alarm probability. When the sensing-time duration reaches a certain value, the increase in the sensing-time duration does not change the false-alarm probability much. Thus, the optimal utilities decrease since less time is used in data transmission. From the curves, it can be seen that the sensing-time duration largely affects the system performance, and thus, it is important to find the optimal sensing-time setting. In Figs. 1–3, we also show the resulted sensing-time duration in our proposed polyblock algorithms and the corresponding utilities, which are marked as "Polyblock Algorithm" in the figures. It is clear that

$$\begin{aligned}
&\max_{\tau,\{\varepsilon_n\},\{p_{n,m}\},\{x_{n,m}\}}\ \min_m\left(\left(1-\frac{\tau}{T}\right)\cdot\sum_{n=1}^{N}\left[\Pr\left(\mathcal{H}_n^0\right)\cdot\left(1-P_n^f(\tau,\varepsilon_n)\right)\cdot r_{n,m}^0+\Pr\left(\mathcal{H}_n^1\right)\left(1-P_n^d(\tau,\varepsilon_n)\right)\cdot r_{n,m}^1\right]\right)\\
&=\max_\tau\ \max_{\{\varepsilon_n\},\{p_{n,m}\},\{x_{n,m}\}}\ \min_m\left(\left(1-\frac{\tau}{T}\right)\cdot\sum_{n=1}^{N}\left[\Pr\left(\mathcal{H}_n^0\right)\cdot\left(1-P_n^f(\tau,\varepsilon_n)\right)\cdot r_{n,m}^0+\Pr\left(\mathcal{H}_n^1\right)\left(1-P_n^d(\tau,\varepsilon_n)\right)\cdot r_{n,m}^1\right]\right)\\
&=\max_\tau\ \left(1-\frac{\tau}{T}\right)\cdot\max_{\{\varepsilon_n\},\{p_{n,m}\},\{x_{n,m}\}}\ \min_m\left(\sum_{n=1}^{N}\left[\Pr\left(\mathcal{H}_n^0\right)\cdot\left(1-P_n^f(\tau,\varepsilon_n)\right)\cdot r_{n,m}^0+\Pr\left(\mathcal{H}_n^1\right)\left(1-P_n^d(\tau,\varepsilon_n)\right)\cdot r_{n,m}^1\right]\right)
\end{aligned}\tag{15}$$

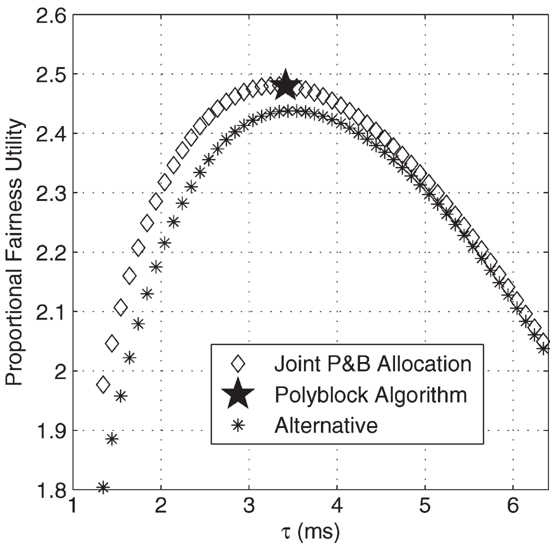

Fig. 2. Proportional fairness utility versus τ.

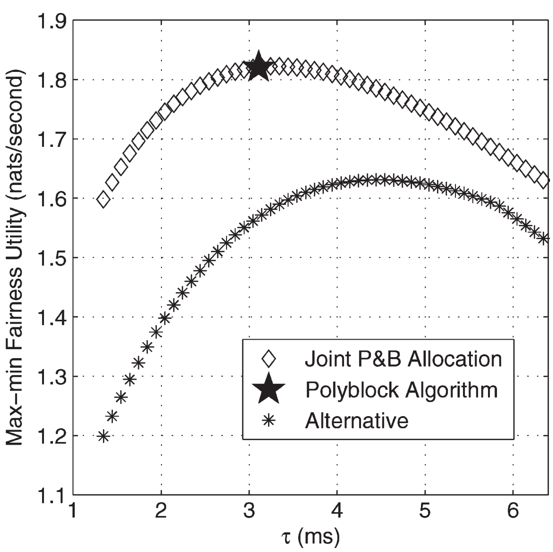

Fig. 3. Max–min fairness utility versus τ.

the utilities obtained in our polyblock algorithms are very close to the global optimal utilities in Figs. 1–3.

Next, we demonstrate the effect of resource allocation in our research, in which each channel can be shared by multiple users. As a comparison, we consider an alternative case when each channel can be used by one secondary user only. The alternative case is equivalent to solving Problems P1, P6, and P9 with additional constraints $x_{n,m} \in \{0, 1\}, m \in \mathcal{M}, n \in \mathcal{N}$. In the newly formulated problems, it is possible that a secondary user is assigned more than one channel, and it is also possible that a secondary user is assigned no channel. It can be proved that the optimal utilities are achieved only when $P_n^d(\tau, \varepsilon_n) = P_{\mathrm{th}}$. Thus, for a given τ, the optimal ε_n can be obtained based on $P_n^d(\tau, \varepsilon_n) = P_{\mathrm{th}}$. Therefore, to maximize the utility for a given τ, we need to determine the optimal power-allocation $(p_{n,m})$ and channel-allocation $(x_{n,m})$ strategies. This is a mixed-integer problem, which is usually NP-hard. To solve the problem, we exhaustively search all the combinations of assigning N channels to M secondary users. In each combination, the power allocation (over probably multiple channels) at each secondary user is determined by a convex optimization problem. Then, for each τ, the optimal power- and channel-allocation strategies are the strategies associated with the combination whose optimal utility has the largest value. For the alternative case, the optimal utilities for different τ are also plotted in Figs. 1–3, which are marked as "Alternative." It can be seen that the alternative case has worse performance than our polyblock algorithms. This is because our algorithms take advantage of dynamic sharing of each channel. Another advantage of our algorithms over the algorithms in the alternative case is the computational complexity. Apparently, our algorithms are significantly less complex since our algorithms are for problems with continuously valued variables, whereas the alternative case is associated with mixed-integer problems.

VI. Conclusion

In this paper, the problem of optimal multichannel cooperative sensing and resource allocation (of spectrum and power) in cognitive radio networks is explored. The issues of how to set spectrum-sensing time, how to determine the spectrum-sensing threshold, how to allocate spectrum resources, and how to set transmission power are investigated. A weighted throughput maximization problem is formulated. Although nonconvex, the problem is solved with the aid of bilevel optimization and monotonic programming methods. We also show that bilevel optimization and monotonic programming methods are applicable when proportional or max–min fairness is considered.

References

[1] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[2] E. Hossain, D. Niyato, and Z. Han, *Dynamic Spectrum Access in Cognitive Radio Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[3] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 589–600, Apr. 2007.

[4] H. Jiang, L. Lai, R. Fan, and H. V. Poor, "Optimal selection of channel sensing order in cognitive radio," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 297–307, Jan. 2009.

[5] R. Fan and H. Jiang, "Channel sensing-order setting in cognitive radio networks: A two-user case," *IEEE Trans. Veh. Technol.*, vol. 58, no. 9, pp. 4997–5008, Nov. 2009.

[6] L. Lai, H. El Gamal, H. Jiang, and H. V. Poor, "Cognitive medium access: Exploration, exploitation and competition," *IEEE Trans. Mobile Comput.*, vol. 10, no. 2, pp. 239–253, Feb. 2011.

[7] A. Motamedi and A. Bahai, "Optimal channel selection for spectrum-agile low-power wireless packet switched networks in unlicensed band," *EURASIP J. Wireless Commun. Netw.*, vol. 2008, pp. 1–10, art. 896420. 2008.

[8] Y.-C. Liang, Y. Zeng, E. C. Y. 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.

[9] Z. Han, R. Fan, and H. Jiang, "Replacement of spectrum sensing in cognitive radio," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2819–2826, Jun. 2009.

[10] T. Wang, C. Li, and H.-H. Chen, "An iterative expectation–maximization algorithm based joint estimation approach for CDMA/OFDM composite radios," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3196–3205, Aug. 2008.

[11] R. Fan and H. Jiang, "Optimal multi-channel cooperative sensing in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 3, pp. 1128–1138, Mar. 2010.

[12] H. Su and X. Zhang, "Cross-layer based opportunistic MAC protocols for QoS provisionings over cognitive radio mobile wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 118–129, Jan. 2008.

[13] J. Jia, Q. Zhang, and X. Shen, "HC-MAC: A hardware-constrained cognitive MAC for efficient spectrum management," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 106–117, Jan. 2008.

[14] A. Ghasemi and E. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *Proc. IEEE Symp. New Frontiers Dyn. Spectr. Access Netw.*, Baltimore, MD, Nov. 2005, pp. 131–136.

[15] X. Kang, Y.-C. Liang, H. K. Garg, and L. Zhang, "Sensing-based spectrum sharing in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4649–4654, Oct. 2009.
[16] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, "Optimal multiband joint detection for spectrum sensing in cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1128–1140, Mar. 2009.
[17] K. B. Letaief and W. Zhang, "Cooperative communications for cognitive radio networks," *Proc. IEEE*, vol. 97, no. 5, pp. 878–893, May 2009.
[18] X. Gong, S. A. Vorobyov, and C. Tellambura, "Joint bandwidth and power allocation with admission control in wireless multi-user networks with and without relaying," *IEEE Trans. Signal Process.*, accepted with minor revision.
[19] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[20] C. Floudas and P. M. Pardalos, *Encyclopedia of Optimization*, 2nd ed. New York: Springer-Verlag, 2009.

A Layered Decomposition Framework for Resource Allocation in Multiuser Communications

Cho Yiu Ng, *Student Member, IEEE*,
Kenneth W. Shum, *Member, IEEE*, Chi Wan Sung, *Member, IEEE*,
and Tat Ming Lok, *Senior Member, IEEE*

***Abstract*—The resource allocation problem for multiuser channels is decomposed into two layers. The lower layer is the weighted sum rate maximization, which is widely considered for many different multiuser channels. The weighted sum rate maximization is employed as a subroutine and called the upper layer. The upper layer is the optimization of dual variables for maximizing a joint utility function, which is very general and includes proportional fairness and max-min fairness as special cases. This layered approach decouples the physical-layer technologies from system-layer consideration. For example, if we want to evaluate a new objective in resource allocation, changes are required only in the outer loop, whereas the inner loop remains the same. This induces flexibility in the software structure. To numerically obtain the solution, a Gauss–Seidel-type algorithm is proposed, and its effectiveness in achieving proportional fairness in the parallel Gaussian broadcast channel is demonstrated by computer simulation.**

***Index Terms*—Dual decomposition, network utility maximization, orthogonal frequency-division multiple-access (OFDMA), proportional fairness.**

I. Introduction

For multiuser communications, a common performance measure is the sum rate of all users. For downlink cellular systems, as the distances between the users and the base station may spread over a very wide range, maximizing the sum rate will penalize users who are far away from the base station. To alleviate this near–far unfairness, one solution is to maximize the weighted sum rate. Weighted sum rate maximization for a parallel Gaussian broadcast channel (BC) is considered in [1] and [2], and that for orthogonal frequency-division multiple-access (OFDMA) systems is considered in [3]. On the other hand, it is unclear how to pick a set of good weighting factors. Another alternative is to assign data rates according to some notion of fairness, e.g., proportional fairness [4], max-min fairness, etc. While the proportional fairness problem for OFDMA systems has been considered in [5], we consider the same problem for the parallel Gaussian BC but under a more general framework based on the concept of joint utility function.

In view of the fact that much research effort has been put into maximizing the achievable rate region of a given multiuser communication channel, our framework employs the weighted sum rate maximization algorithm, which is a common way to characterize the rate region, as a subroutine. This isolates the physical-layer issue from other higher layer consideration. It can be applied to many multiuser channel models such as the parallel Gaussian BC, fading multiple access channel [6], and the cooperative transmission scheme in [7]. Whenever there is a need to adopt a new resource-allocation objective or an advance in physical-layer technologies, we only need to change some of the modules rather than redevise the whole optimization algorithm. Such flexible software structure can reduce the development cost and maintenance cost of the resource-allocation software. In addition, it also allows researchers to decouple a complex problem and focus on one of the aspects.

The optimization framework discussed in this paper is based on a layered decomposition approach similar to those in [8] and [9]. However, a more general joint utility function is adopted in this paper, whereas, in [8] and [9], the utility function to be optimized is assumed to be separable, i.e., it is the sum of the utility functions pertaining to the users. The approach in this paper is thus more flexible. Our iterative algorithm, which updates the dual variables in a Gauss–Seidel manner, is more efficient than the subgradient approach suggested in [8] and [9]. To illustrate the difference in computation time, we compare our algorithm with an algorithm proposed in [10], which is based on the subgradient approach and designed for the parallel Gaussian BC. Simulation results show that our algorithm has faster convergence. On the other hand, we remark that the subgradient approach is more appropriate for distributed implementation. Our proposed algorithm in this paper can be considered as a tradeoff between performance and the how distributed the algorithm is.

In Section II, we present a joint utility optimization for parallel Gaussian BC as a motivating example. The decomposition framework and convergence results are detailed in Section III. Some numerical examples are given in Section IV.

II. Motivating Example and Problem Formulation

Consider a family of K parallel Gaussian BCs with N receivers. For $n = 1, \ldots, N$ and $j = 1, \ldots, K$, let $g_{n,j}$ be the power gain for $j = 1, 2, \ldots K$ user n in channel j, and let π_j be a permutation over the set $\{1, 2, \ldots, N\}$ such that $g_{\pi_j(i_1),j} > g_{\pi_j(i_2),j}$ if $i_1 < i_2$. According to [1], a rate vector $\mathbf{R} \triangleq (R_1, \ldots, R_N)$ is achievable in the parallel Gaussian BC if we can write

$$P = \sum_{n=1}^{N} \sum_{j=1}^{K} p_{n,j} \quad \text{and} \quad R_n = \sum_{j=1}^{K} R_{n,j}$$

Manuscript received October 20, 2009; revised March 24, 2010, July 19, 2010, and September 27, 2010; accepted November 8, 2010. Date of publication December 3, 2010; date of current version February 18, 2011. This work was supported in part by the grants from the Research Grant Council of the Hong Kong Special Administrative Region, China under Project CityU 120107 and Project CUHK 415608. The review of this paper was coordinated by Prof. Y. Su.

C. Y. Ng and T. M. Lok are with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: michaelng@ieee.org; tmlok@ie.cuhk.edu.hk).

K. W. Shum is with the Institute of Network Coding, The Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: wkshum@inc.cuhk.edu.hk).

C. W. Sung is with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong (e-mail: albert.sung@cityu.edu.hk).

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/TVT.2010.2096439