This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2010 proceedings
This paper was presented as part of the main Technical Program at IEEE INFOCOM 2010.

Asymmetry-Aware Real-Time Distributed Joint Resource Allocation in IEEE 802.22 WRANs

Hyoil Kim and Kang G. Shin
Department of Electrical Engineering and Computer Science
The University of Michigan, Ann Arbor, MI 48109–2121
Email: {hyoilkim, kgshin}@eecs.umich.edu

***Abstract*—In IEEE 802.22 Wireless Regional Area Networks (WRANs), each Base Station (BS) solves a complex resource allocation problem of simultaneously determining the channel to reuse, power for adaptive coverage, and Consumer Premise Equipments (CPEs) to associate with, while maximizing the total downstream capacity of CPEs. Although joint power and channel allocation is a classical problem, resource allocation in WRANs faces two unique challenges that has not yet been addressed: (1) the presence of small-scale incumbents such as wireless microphones (WMs), and (2) asymmetric interference patterns between BSs using omnidirectional antennas and CPEs using directional antennas. In this paper, we capture this asymmetry in upstream/downstream communications to propose an accurate and realistic WRAN-WM coexistence model that increases spatial reuse of TV spectrum while protecting small-scale incumbents. Based on the proposed model, we formulate the resource-allocation problem as a mixed-integer nonlinear programming (MINLP) which is NP-hard. To solve the problem in real-time, we propose a suboptimal algorithm based on the Genetic Algorithm (GA), and extend the basic GA algorithm to a fully-distributed GA algorithm (dGA) that distributes computational cost over the network and achieves scalability via local cooperation between neighboring BSs. Using extensive simulation, the proposed dGA is shown to perform as good as 99.4–99.8% of the optimal solution, while reducing the computational cost significantly.**

I. INTRODUCTION

Cognitive radio (CR) is a key technology for alleviating the inefficient spectrum-utilization problem under the current static spectrum-allocation policy [1]. In CR networks (CRNs), unlicensed or secondary users (SUs) are allowed to *opportunistically reuse* spectrum bands assigned to licensed or primary users (PUs) as long as they do not cause any harmful interference to the PUs.

Resource allocation is one of the key challenges in CRNs. Resources in CRNs—as in traditional wireless networks—include channel and power of each radio. However, how to allocate such resources in CRNs is more difficult than in traditional wireless networks, primarily because PUs' use of licensed channels must be protected while allowing SUs to opportunistically use them. For this, the concept of interference temperature has been proposed, specifying the maximum interference PUs can tolerate. Therefore, CRs must carefully choose their channel and transmit power so as not to interfere with PUs more than allowed.

In this paper, we study resource allocation in IEEE 802.22 Wireless Regional Area Networks (WRANs), the first international CRN standard that reuses VHF/UHF TV bands.[1] A 802.22 WRAN is a cellular network consisting of (1) a base station (BS) that reaches up to 100 km radius and (2) CR end-users associated with the BS within its coverage, referred to as Consumer Premise Equipments (CPEs). The CPEs represent households in a rural area and thus they are static nodes.

A. Motivation

Resource allocation in IEEE 802.22 introduces two unique challenges. First, it is a three-dimensional problem of making decisions on how to assign *channel* and *power* to a BS, and which BS a CPE should be associated with (called *BS-CPE association*). The BS-CPE association is considered because IEEE 802.22 supports *self-coexistence* where multiple BSs may co-exist in the same region with different coverage. Second, resource allocation must account for coexistence with the incumbents (i.e., PUs) in TV bands: analog TV (ATV), digital TV (DTV), and wireless microphones (WMs). Therefore, a good resource allocation algorithm should *not only* consider efficiency of spectrum reuse *but also* ensure protection of PUs.

To address such challenges, we develop an accurate WRAN-PU coexistence model by focusing on the asymmetry between BS-to-PU interference and CPE-to-PU interference, rooted at the realism that a BS and CPEs are using different types of antenna. In 802.22, a BS is equipped with an omnidirectional antenna for downlink communication with CPEs, generating an isotropic BS-to-PU interference pattern. By contrast, each CPE is equipped with a directional antenna for its uplink communication with the BS, thus generating a directional CPE-to-PU interference pattern. Nevertheless, resource allocation in the CR literature has only considered the case where each network entity is equipped solely with an omnidirectional antenna. Hence, we must look at the problem from a new angle by considering this asymmetry.

Among the three types of PUs in TV bands (i.e, ATV, DTV, and WMs), we consider WMs as a major target to coexist with. Although our work can be extended to address ATV and DTV users, coexistence with WMs is the most challenging and has the most practical value for the following reasons. First, after the scheduled ATV-to-DTV transition in 2009, ATV users will not be our major concern. Second, after the transition, some TV channels will be free of TV users, thanks to the efficient channel utilization of DTV technology [2], so such

[1] Note that 802.22 is still a draft.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2010 proceedings
This paper was presented as part of the main Technical Program at IEEE INFOCOM 2010.

DTV-free channels might be used in the early deployment of IEEE 802.22 before expanding its operation to full TV bands. In such a case, coexistence with WMs becomes a major issue. Finally, coexistence with small-scale PUs like WMs is a more serious issue, because spatial variation of spectrum availability caused by them is more fine-grained: the transmission range of a WM is only up to 100 m, whereas the keep-out radius from a DTV transmitter is 150.3 km [3]. Therefore, in most cases BSs will operate outside of the DTV keep-out region but they still have to deal with WMs which can appear at any location within their cells. In spite of the short communication range, even a single WM can force a large 802.22 cell to vacate the channel it resides at, which deteriorates efficiency in spatial spectrum reuse significantly.

Despite its importance, resource allocation with small-scale incumbents (i.e., WMs) is not yet studied thoroughly. Hence, we will focus on the problem of three dimensional resource allocation—channel, power, and BS-CPE association—while accounting for the impact of asymmetric interference patterns between a BS and CPEs. This way, we would like to establish an important stepping stone towards better spatial reuse of TV bands in IEEE 802.22.

B. Contributions

Our contribution in this paper is three-fold. First, we propose a realistic WRAN-WM coexistence model that captures asymmetry in interference patterns between BS-to-WM and CPE-to-WM cases. The proposed model is presented in Section IV and will serve as a basis for our problem formulation.

Next, we provide a mathematical formulation for finding an optimal solution to the joint channel & power allocation and BS-CPE association, based on the key conditions for WM protection. The problem turns out to be mixed-integer nonlinear programming (MINLP), which is known to be NP-hard. Its details are presented in Section V.

Finally, to solve the NP-hard problem, we propose a suboptimal algorithm by applying the Genetic Algorithm (GA). The use of GA is justified by the need to have a *real-time* solvable algorithm with reasonable performance, since IEEE 802.22 requires a WRAN to vacate its channel within 2 seconds from the appearance of an in-band incumbent – implying that resource re-allocation should also be performed until the same deadline. We first develop the basic GA with its key components, and then show both centralized GA (cGA) and parallel GA do not scale well with the network size, i.e., the number of BSs. Therefore, we extend the standard GA to a fully distributed GA (dGA), tailored to a large cellular networks with hexagonal BS deployment. The proposed algorithm is presented in Section VI.

C. Organization

We first overview the related work in Section II. After Section III briefly reviews IEEE 802.22 and our assumptions, Section IV presents our proposed WRAN-WM coexistence model. In Section V, we provide an analytic basis of the joint resource allocation problem and cast it into an MINLP. Section VI describes formulation of the basic GA and then presents our proposed dGA. The performance of the proposed algorithm is evaluated and compared with the global optimum as well as the performance of cGA in Section VII. The paper concludes with Section VIII.

II. Related Work

In recent years, joint channel and power allocation has been studied in the context of CRNs, and we briefly review the literature. Digham [4] studied uplink power assignment and channel allocation to optimize uplink capacity of CPEs in a single cell, and Che *et al.* [5] investigated a similar problem with emphasis on Quality-of-Service (QoS) provisioning for CRs. However, both work ignored BS-to-PU interference. He *et al.* [6] and Li *et al.* [7] studied joint channel and power allocation in an ad-hoc CRN. However, the ad-hoc scenario does not capture the asymmetry of CR-to-PU interference in uplink (by CPEs) and downlink (by BS) communications. Unlike the above-mentioned related work, we deal with joint allocation of channel, power, and BS-CPE association, and discuss coexistence of small-scale incumbents with CPEs that have directional transmission patterns.

Except some special cases [4] where convex optimization is applicable, joint resource allocation is generally NP-hard. The Genetic Algorithm (GA) is a suboptimal approach to solving this type of problem efficiently [8], [9], and its potential for CRNs has been shown by ElNainay *et al.* [9] who applied the parallel GA to a joint channel and power allocation problem. The parallel GA, however, still has limitations since it does not fully scale with a network size. Therefore, we develop a distributed GA mechanism that requires only local cooperation between neighboring BSs for better scalability.

III. Preliminaries

In this section, we briefly overview the IEEE 802.22 WRAN and introduce the assumptions used in the paper.

A. IEEE 802.22—An Overview

An IEEE 802.22 WRAN is a cellular network where a BS coordinates CPEs with coverage up to 100 km (typically 33 km). A CPE represents a household (thus stationary) in a rural area. The WRAN reuses TV bands, 6 MHz each in the US, by utilizing them based on OFDMA modulation for downstream and upstream links [10].

Each BS employs an omnidirectional antenna for downstream communications, and CPEs use directional antennas for their upstream communications. The use of directional antennas at CPEs is crucial in IEEE 802.22: if CPEs are using omnidirectional antenna instead, CPEs at fringe areas have to use high power to reach the BS, say R meters away, doubling the interference range of a WRAN from R to $2 \cdot R$. Directional antennas, however, can concentrate the power towards the BS while minimizing the backward signal emission. For example, 802.22 limits the directional antenna's half-power beamwidth to 60 degrees [11] and back-to-front ratio to 16 dB [12].

B. Assumptions

When a WM emerges, CPEs nearby the WM may need to re-associate themselves with different BSs to avoid interference at the WM. For this, a procedure called *antenna azimuth adjustment* is defined in the draft to describe how to make a directional antenna properly oriented towards the selected BS [13]. However, it does not specify whether the azimuth adjustment is performed manually or automatically.

In this paper, we assume that each CPE is equipped with a directional antenna that is automatically-adjustable in real-time, with the following rationale. First, if azimuth adjustment is done manually, it can be performed only once at the installation of an antenna implying a CPE's home BS should be fixed.[2] This introduces a huge disadvantage in resource utilization because when a WM appears, a nearby CPE may only have two choices: (1) to turn off its power until the WM disappears, or (2) to make its home BS and other CPEs in the same cell to move into another channel. Clearly, both options are not desirable in terms of fair and efficient spectrum reuse. Second, such flexible antennas are already within our reach for practical usage. For example, Ruckus Wireless introduced the *BeamFlex* Smart Wi-Fi antenna [14], that is able to reconfigure its directional antenna elements in real time. BeamFlex-equipped access points are as low as \$117, which is comparable to other off-the-shelf 802.11 APs. Although WRANs require a longer range, we expect to have cheaper antennas in the near future thanks to the cost reduction of a new technology after its launching to the market driven by the market competition.

We also assume that locations of all WMs, BSs, and CPEs are known *a priori*, which is reasonable because (1) locations of WMs are revealed by IEEE 802.22.1 beaconing via the "Location" field in the MPDU of the beacon frame [15], and (2) locations of BSs and CPEs can be looked up from the *geolocation database* [16] which is a mandated feature in the 802.22 standard. For 802.22.1 beacons, we need to deploy spectrum sensors with a desired density while considering the fact that the 802.22.1 beacon ranges up to 3 km, e.g., at least one sensor within a 3 km radius from a WM. Each sensor reports its detection results to the BS so that the location of the detected WM is revealed. The issues in WM detection and sensor deployment have already been discussed in the literature, and thus, we do not discuss them any further in this paper. Interested readers may refer to [17] for WM detection and [18] for sensor density.

IV. Proposed Model for WRAN-WM Coexistence

Coexistence of CRs with PUs is essential to success of WRANs. For this, the impact of CR-to-PU interference must be accounted for by considering two types of PUs, large-scale PUs (e.g., DTV transmitters) and small-scale PUs (e.g., WMs). Protection of large-scale PUs has been studied extensively with focus on determining whether a WRAN is within the keep-out radius of a DTV transmitter [3]. Coexistence with small-scale PUs, however, becomes far more complex due to their small-scale footprints. For example, even one WM with small transmission power may force a whole WRAN to move out of the WM's channel. Therefore, we focus on WRAN-WM coexistence to achieve better spatial reuse of spectrum.

We first argue that the coexistence guideline in the current draft is too strict and thus less efficient. When an incumbent is operating in channel k, the draft recommends that CRs may not use the same channel k or adjacent channels k+/-1 to avoid CR-to-PU interference. Although this recommendation is proper for wide-band PUs such as DTV transmitters, it is not efficient for coexistence with narrow-band WM signals occupying just $1/30$ (= 200 KHz / 6 MHz) of a TV band. In fact, it has been shown that CRs can still coexist with the WMs in adjacent channels using a guard band [20]. We will further show that CRs can also use the same channel with WMs as long as the interference constraints in the draft are not violated, and then we propose an accurate and realistic model for WRAN-WM coexistence.

Before introducing the proposed model, we need to understand asymmetry in WRAN-to-WM interference in terms of interference temperature. We first consider BS-to-WM interference, where a BS has a coverage area of A_0 of radius R_0 that is determined by the minimum field strength of $E_0 = 28.8$ dB(uV/m),[3] as shown in Fig. 1. Since a BS is using omnidirectional antenna, its pollution area becomes circular as illustrated in the figure by A_1 for co-channel WMs and by A_3 for adjacent-channel WMs. The radius R_1 of A_1 is derived by the maximum allowed interference field strength at a WM: $E_1 = 32.7$ dB(uV/m) [19]. For A_3, out-of-band emission mask of 33 dB is assumed, i.e., a BS's signal is observed with 33 dB less field strength at its adjacent channels, and thus R_3 is derived by the field strength of $E_3 = 32.7 + 33$ dB(uV/m).[4] This implies that even when a co- or adjacent-channel WM appears inside the coverage of a BS (i.e., A_0), the BS may be able to use the channel unless the WM is inside A_1 or A_3.

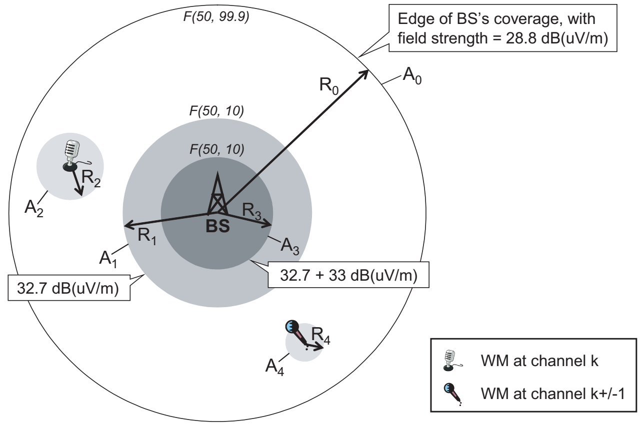
The CPE-to-WM interference is also an important factor, which is described by the interference range of a CPE, denoted by A_2 and A_4 for co- and adjacent-channel coexistence, respectively. However, the previous model in [19], presented in Fig. 1(a), did not capture the impact of using directional antennas by CPEs that makes the interference pattern "directional", not circular. That is, even though a CPE is located outside A_2 and A_4 of Fig. 1(a), it can still interfere with the WM if the WM is located on the line of sight between the CPE and the BS. Therefore, A_2 and A_4 must be re-shaped properly to protect WMs against interference from CPEs.

To correctly model CPE-to-WM interference, we propose a new WRAN-WM coexistence model in Fig. 1(b). In this model, A_2 and A_4 are now fan-shaped with widths denoted by Θ_1 and Θ_2, which are chosen wide enough to ensure that the

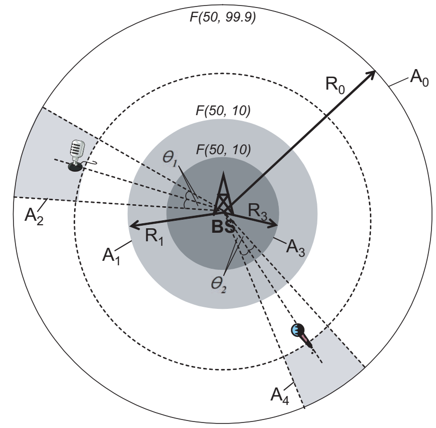
[2] In this paper, the *home BS* implies the BS a CPE is associated with, and other BSs not associated with the CPE are called *foreign BSs*.

[3] such that the measured field strength at R_0 exceeds E_0 at 50% of the locations for 99.9% of time, i.e., $F(50, 99.9)$.

[4] For R_1 and R_3, $F(50, 10)$ is used for maximal protection of WMs.



(a) A previous model in [19]



(b) The new model we propose

Fig. 1. CR-WM coexistence model

WM would not be interfered with by the directional antenna's emission pattern spreading over 60 degrees. Determination of Θ_1 and Θ_2 depends on the 'guard band' of the scheme in [20], which is not within the scope of this paper. In Section V, we will use this new model to propose asymmetric joint resource allocation with both BS and CPEs considered.

V. Optimal Resource Allocation in WRANs

In this section, we study how to maximize the sum rate of all CPEs considering their downstream communications, while satisfying the PU protection requirements by controlling interference due to the transmissions from BSs and CPEs. There are two reasons for focusing on downstream communications. First, the CPEs in the overlapping area of two or more BSs may experience degraded SINR due to interference from foreign BSs. Second, the CPE-to-'foreign BS' interference in upstream communications is efficiently minimized by using the directional antennas and thus is of less interest.

We consider the problem where each BS seeks at most one channel to operate in, since the current draft does not include channel bonding as an option [21]. We leave multi-channel allocation as our future work, in case the future standard includes the channel boding concept.

We assume there are L WMs, M BSs, and N CPEs, each operating in one of K TV channels. In addition, we use $l \in \{1, 2, \dots, L\}$, $m \in \{1, 2, \dots, M\}$, $n \in \{1, 2, \dots, N\}$, and $k \in \{1, 2, \dots, K\}$ as indices to identify each WM, BS, CPE, and channel, respectively. We also use $w_l \in \{1, 2, \dots, K\}$ to denote the channel used by WM l. We further denote by D^{BW}_{ml} the Euclidean distance between BS m and WM l, by D^{BC}_{mn} the Euclidean distance between BS m and CPE n, and by D^{CW}_{nl} the Euclidean distance between CPE n and WM l.

A. Power & Channel Allocation of a BS

We denote by P_{mk} the power of BS m on channel k, whose range is given as

$$P^{min}_{mk} \cdot s_{mk} \le P_{mk} \le P^{max}_{mk} \cdot s_{mk}, \tag{1}$$

where P^{min}_{mk} is the minimum power of BS m to have at least one CPE within its coverage A_0, P^{max}_{mk} is the maximum power BS m is allowed to use so as not to interfere with any WMs by the BS's transmission, and s_{mk} is an indicator for BS m's channel allocation such that

$$s_{mk} = \begin{cases} 1, & \text{if channel } k \text{ is assigned to BS } m \text{ ,} \\ 0, & \text{otherwise ,} \end{cases} \tag{2}$$

which should satisfy

$$\sum_k s_{mk} \le 1, \tag{3}$$

since each BS only uses at most one channel. Here $\sum_k s_{mk} = 0$ implies the BS m is completely turned off. In addition, P_{mk} must be zero in those channels not assigned to the BS (i.e., $s_{mk} = 0$), which is satisfied by the term s_{mk} in Eq. (1).

For a given transmitted power P_{mk} of BS m, R_0, R_1, and R_3 are calculated according to the procedures in the ITU-R P.1546-3 propagation model [22] and the WRAN reference model [23]. Those models utilize actual measurements on land to provide signal attenuation plots as a function of distance from the transmitter and conversion tables between transmitted EIRP (or ERP) and the field strength at a given distance. Interested readers may refer to the references for more details on the procedure.

For ease of notation, the radii of A_0, A_1, and A_3 will be denoted by $R_0(P_{mk})$, $R_1(P_{mk})$, and $R_3(P_{mk})$, provided BS m operates in channel k with power P_{mk}.

1) Determination of P^{max}_{mk}: P^{max}_{mk} is a function of A_1 and A_3 such that at any $P_{mk} \le P^{max}_{mk}$, the closest WM at channel k may not be inside A_1 and at the same time, the closest WM at channel $k \pm 1$ may not be inside A_3. Therefore, we have

$$\begin{aligned} P^{max}_{mk} &= \max\{P_{mk} | R_1(P_{mk}) \le \bar{R}_1 \text{ and } R_3(P_{mk}) \le \bar{R}_3\}, \\ \bar{R}_1 &= \begin{cases} min_{\{l|w_l=k\}}\{D^{BW}_{ml}\}, & \text{if } \{l|w_l = k\} \ne \emptyset, \\ \infty, & \text{otherwise,} \end{cases} \\ \bar{R}_3 &= \begin{cases} min_{\{l|w_l=k\pm1\}}\{D^{BW}_{ml}\}, & \text{if } \{l|w_l = k \pm 1\} \ne \emptyset, \\ \infty, & \text{otherwise.} \end{cases} \end{aligned}$$

2) Determination of P^{min}_{mk}: P^{min}_{mk} is determined by the closest CPE from BS m, since P_{mk} has to be large enough to include at least one CPE within its coverage of radius R_0;

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2010 proceedings
This paper was presented as part of the main Technical Program at IEEE INFOCOM 2010.

otherwise, the BS would become a pure interferer. Therefore, we have

$$P_{mk}^{min} = \{P_{mk} | R_0(P_{mk}) = min_n\{D_{mn}^{BC}\}\}.$$

Note that in case $P_{mk}^{min} > P_{mk}^{max}$, channel k should not be used by BS m by setting $s_{mk} = 0$.

B. CPE-BS Association

A CPE can associate itself with at most a single BS. We use t_{nm} as an indicator for CPE n's association with a BS such that

$$t_{nm} = \begin{cases} 1, & \text{if CPE } n \text{ is associated with BS } m\ , \\ 0, & \text{otherwise}\ , \end{cases} \quad (4)$$

which should satisfy

$$\sum_m t_{nm} \leq 1, \quad (5)$$

that is, a CPE may not associate itself with any BS (i.e., $t_{nm} = 0, \forall m$) if it is not in the coverage of any BS.

There are three conditions forcing $t_{nm} = 0$ as follows:

C1: $D_{mn}^{BC} > \sum_k R_0(P_{mk}) s_{mk}$,

C2: $D_{ml}^{BW} \leq D_{mn}^{BC}$ & $\cos^{-1}\left(\frac{[D_{mn}^{BC}]^2 + [D_{ml}^{BW}]^2 - [D_{nl}^{CW}]^2}{2 \cdot D_{mn}^{BC} \cdot D_{ml}^{BW}}\right) \leq \Theta_1$ for (k, l) such that $s_{mk} = 1$ and $w_l = k$,

C3: $D_{ml}^{BW} \leq D_{mn}^{BC}$ & $\cos^{-1}\left(\frac{[D_{mn}^{BC}]^2 + [D_{ml}^{BW}]^2 - [D_{nl}^{CW}]^2}{2 \cdot D_{mn}^{BC} \cdot D_{ml}^{BW}}\right) \leq \Theta_2$ for (k, l) such that $s_{mk} = 1$ and $w_l = k \pm 1$.

The above conditions imply that, for association with BS m, (**C1**) CPE n must be within A_0 of BS m, (**C2**) CPE n must *not* be within A_2 of BS m using channel k for any l such that $w_l = k$, and (**C3**) CPE n must *not* be within A_4 of BS m using channel k for any l such that $w_l = k \pm 1$. Note that in **C2** and **C3**, the law of cosines was applied.

C. Objective Function

We want to maximize the sum rate of CPEs by considering SINR for each CPE-BS pair. That is,

$$\textbf{Maximize} \sum_n \sum_m \sum_k t_{nm} s_{mk} B \cdot log_2(1 + SINR_{nmk}), \quad (6)$$

subject to Eqs. (1)-(5) and **C1-C3**

where $SINR_{nmk}$ is the SINR of CPE n associated with BS m using channel k, expressed as

$$SINR_{nmk} = \frac{P_{mk} H_{mn}}{\sum_{m' \neq m} s_{m'k} P_{m'k} H_{m'n} + N_0 B},$$

where H_{mn} is the path gain between BS m and CPE n which is derived from the ITU-R P.1546-3 propagation model, and $N_0 = -163$ dBm/Hz is the noise power spectral density (PSD) [3] and $B = 6$ MHz. Note that since the ITU-R model is built on actual field measurements, this approach results in more realistic resource allocation than assuming exponential decaying path-loss such as $H_{m,n} = (D_{mn}^{BC})^{-\alpha}$.

The form of Eq. (6) indicates that the problem is a mixed-integer nonlinear programming (MINLP) which is NP-hard in general. Although one may be able to use integer relaxation or LP approximation techniques to derive polynomial-time solvable suboptimal algorithms, such algorithms do not satisfy the PU protection requirements in IEEE 802.22. The draft indicates that a WRAN should vacate a channel *within* 2 *seconds* upon detection of an emerging incumbent in the same channel. Since the appearance of new PUs triggers re-allocation of spectrum resources (i.e., power and channel), we can conclude that we need a *real-time* solvable algorithm completing in 2 seconds. Therefore, in the next section, we propose a suboptimal resource allocation algorithm using a genetic algorithm (GA) that can control when to terminate the algorithm, and extend it to a distributed version for scalability with multiple WRAN cells.

VI. Suboptimal Resource Allocation via Distributed Genetic Algorithm

Genetic Algorithms (GAs) [24], [25] are considered as an efficient suboptimal approach to an NP-hard problem. GAs are based on Darwin's theory of evolution where each possible solution of the optimization problem is modeled as an *individual* of a species that is characterized by its chromosome. A pool of such individuals is considered as a *generation* or *genetic pool* of the species in its evolutionary chain.

The GA evolves its generation into the next generation via the three essential steps: selection, crossover/mating, and mutation. That is, the individuals of the next generation are created by (1) selecting genetically superior individuals as parents from the current genetic pool, (2) mating them to produce their offsprings, and (3) occasionally mutating the chromosome of the offsprings to reduce the chance of falling into local optima. Here, the superiority of an individual is measured by 'fitness', which is usually a function of the network objective to optimize.

Via the iterative process of evolution, the GA gradually enhances its gene pool closer to the optimal solution. Although there is no absolute guarantee that the algorithm converges to a global optimum, in many applications it has been shown that the GA can achieve near-optimal performance [9]. One of the key advantages in using the GA is that we can control when to terminate the algorithm by stopping the evolution at any time needed and using the best individual in the current generation.

A. The Basic GA—A Centralized Approach

We first describe formulation of a standard GA and its basic components, and then present the proposed distributed GA in Section VI-B.

1) Definition of Chromosome: We define chromosome as

$$\mathbf{C} = \{\mathbf{b}, \mathbf{PL}, \mathbf{T}\},$$

where $\mathbf{b} = \{b_1, b_2, \ldots, b_M\}$, $\mathbf{T} = \{T_1, T_2, \ldots, T_N\}$, and $\mathbf{PL} = \{PL_1, PL_2, \ldots, PL_M\}$. There are three variables to optimize: BS m's channel denoted by $b_m \in \{1, 2, \ldots, K\}$, BS m's discretized power level denoted by $PL_m \in \{1, 2, \ldots, PL^{max}\}$ ($PL_m \in \mathbb{Z}$), CPE n's home BS index denoted by $T_n \in \{1, 2, \ldots, M\}$.

PL_m discretizes the power of BS m according to the radius R_0 of its coverage. For example, suppose $R_0 \in \mathbf{R} =$

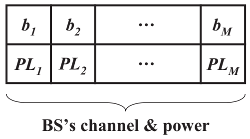


Fig. 2. An illustration of chromosome

$\{0, 10, \ldots, 70\}$ (in km) such that $PL_m = 1$ corresponds to $R_0 = 0$ and $PL_m = PL^{max} = 8$ corresponds to $R_0 = 70$. Then $PL_m = i$ implies the power $\{P_{mk} | R_0(P_{mk}) = \mathbf{R}(i)\}$. PL_m should also satisfy $PL_m \in [PL_{mk}^{min}, PL_{mk}^{max}]$, when $b_m = k$, to conform to the condition in Eq.(1). Here PL_{mk}^{min} and PL_{mk}^{max} are determined as

$$\begin{aligned} PL_{mk}^{min} &= \min\{i | R_0(P_{mk}^{min}) \leq \mathbf{R}(i)\}, \\ PL_{mk}^{max} &= \max\{i | \mathbf{R}(i) \leq R_0(P_{mk}^{max})\}. \end{aligned}$$

This definition of chromosome, however, is inefficient in the sense that the number of possible combinations of b_m, PL_m, and T_n is very large. Each BS would choose one of K channels and PL^{max} power levels, resulting in the total of $K^M \cdot (PL^{max})^M$ combinations. Moreover, each CPE can choose one of M BSs as its home BS, giving M^N possibilities. As a result, the size of the overall search space becomes $O((K \cdot PL^{max})^M \cdot M^N)$.

Therefore, we simplify the definition as in Fig. 2 by choosing T_n's deterministically based on the following observation: if b_m's and PL_m's are given, the best home BS for CPE n is clearly the one giving the highest SINR among the BSs that do not violate **C1–C3** in Section V. In this case, each CPE requires to check at most M BSs, and hence, the total computational cost is $O(M \cdot N)$, which significantly reduces complexity of $O(M^N)$ in the previous definition.

2) Fitness Function: The *fitness* of an individual with chromosome $\mathbf{C}$ is denoted by $f(\mathbf{C})$ and defined as the *normalized sum-rate* by dividing Eq. (6) by B.

3) Algorithm Description: Initially, the GA forms its first generation consisting of Q individuals by randomly generating their chromosomes. When generating $\mathbf{PL}$, the random value of PL_m should be chosen in $[PL_{mk}^{min}, PL_{mk}^{max}]$ (assuming $b_m = k$). The thus-formed generation $\mathbf{G}$ is then described as

$$\mathbf{G} = \{\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_Q\}.$$

Then, the algorithm proceeds by iteratively producing the next generation via the following two stages:

Stage 1 (Cloning): Some superior individuals (in terms of their fitness) are copied to the next generation, to make the maximum fitness of a generation a non-decreasing function. The ratio of the population to be cloned is denoted by *KEEP_RATE* $\in (0, 1)$.

Stage 2 (Breeding): After cloning, there are $Q \cdot (1 -$ *KEEP_RATE*$)$ individuals to fill up in the next generation. These individuals will be produced by the breeding process with the three steps: selection, crossover, and mutation. Each run of the breeding process will create two offsprings, and it will run until the size of the new population reaches Q. The description of the three steps is given as follows.

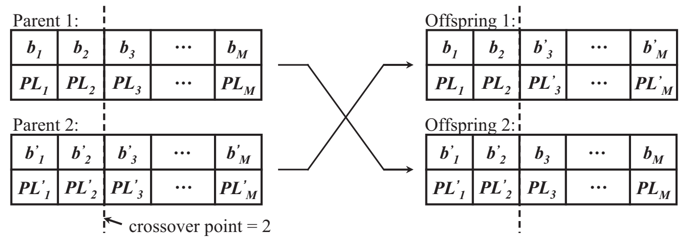


Fig. 3. An example crossover process

- **Selection:** *Selection* chooses two parents for mating. To entice superior individuals, *proportional selection* is employed where an individual j is selected with probability p_j, which is proportional to its fitness such as

$$p_j = f(\mathbf{C}_j) / \sum_{1 \leq j \leq Q} f(\mathbf{C}_j).$$

- **Crossover:** The chosen individuals are mated to produce two offsprings via *1-point crossover*, as shown in Fig. 3. The crossover point is determined randomly between 1 and $(M - 1)$.
- **Mutation:** A new-born offspring is mutated with probability *MUTATION_RATE*, where a single column of its chromosome is chosen for mutation. It then randomly decides whether to mutate PL_m only or (b_m, PL_m) together, and replaces them by randomly generated values.

The algorithm completes when the runtime reaches a pre-defined threshold, i.e., 2 seconds.

B. Distributed Genetic Algorithm (dGA)

In reality, the centralized GA is not practical, since (1) it is a centralized algorithm where each BS should perform its GA individually without cooperation with others, and (2) its computational overhead does not scale with the network size M. Despite the enhancement made in the definition of chromosome, the size of the search space is still $O((K \cdot PL^{max})^M \cdot M \cdot N)$ that may grow exponentially as M increases linearly.

Although parallel GAs may be used to distribute the computational cost over the network [9], it still does not scale well with M. The parallel GAs divide the whole population into M sub-populations so that each sub-population can be searched for a local optimum by a single BS. With this approach, however, the overall search space reduces only by a factor of M resulting in $O((K \cdot PL^{max})^M \cdot N)$. As a result, both basic and parallel GAs cannot be used for real-time resource allocation in the 802.22 WRAN.

Therefore, we propose a fully distributed GA (dGA), tailored to a large cellular network with many cells like in 802.22. Our idea is to upper-bound the computational cost of each BS by making it consider only neighboring BSs and nearby CPEs in its local area.

Assuming hexagonal BS deployment as in Fig. 4, a BS would have at most six neighbor BSs. Then we re-define chromosome as a *sub-chromosome* that only includes seven pairs of $(b_m,\ PL_m)$ for itself (called *center BS*) and six

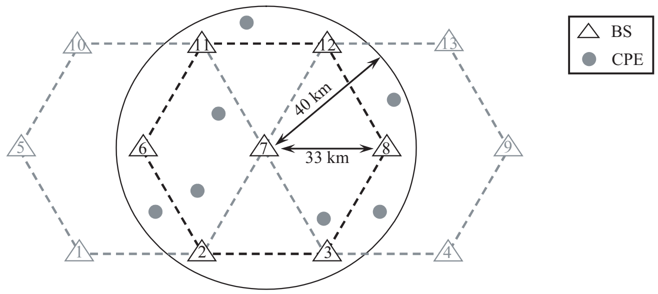


Fig. 4. An illustration of hexagonal BS deployment for dGA, from the view of BS 7 (the number in a BS is its BS index)

BS 7's sub-chromosome:

b_2	b_3	b_6	b_7	b_8	b_{11}	b_{12}
PL_2	PL_3	PL_6	PL_7	PL_8	PL_{11}	PL_{12}

Fig. 5. An example of sub-chromosome: BS 7 in Fig. 4

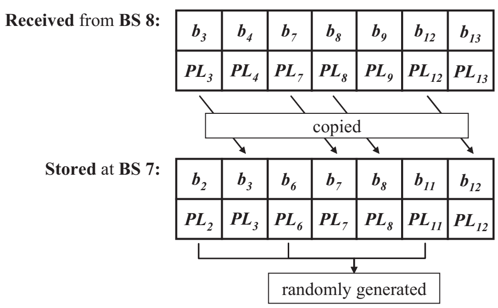


Fig. 6. The process of importing a sub-chromosome from a neighbor BS: when BS 7 receives BS 8's broadcast message in Fig. 4

neighbor BSs. An example of sub-chromosome is illustrated in Fig. 5 for BS 7 in Fig. 4. In addition, the fitness of an individual is calculated by considering the center BS, its six neighbor BSs, and the CPEs within the localized area, which is a circle with radius of *LOCAL_RADIUS* centered at the center BS. This new fitness function is called *local* fitness. For example, in Fig. 4, inter-BS distance is 33 km and the localized area has radius of *LOCAL_RADIUS* = 40 km.

The dGA proceeds as follows. A BS performs a local evolution process where the individuals in the population are defined by the sub-chromosome and their superiority is measured by the local fitness. Once every *EXCHANGE_PERIOD* iterations, the BS picks its best $Q/7$ individuals (assuming Q is a multiple of 7) and broadcasts their sub-chromosomes to the six neighboring BSs. Then, a BS would transmit one broadcast message and listen to six broadcast messages from its neighbors. Using the six sets of $Q/7$ individuals received from the neighbors, the BS reconstructs its current generation such that the new generation will have top $Q/7$ individuals of its own and $6 \cdot (Q/7)$ individuals from the neighbors. Since the imported sub-chromosome from a neighbor BS has at most four common columns, a receiving BS should replace the uncommon columns with its own necessary indexes by randomly generating their channels and power levels. This process is illustrated in Fig. 6.

The dGA completes when the runtime reaches 2 seconds, similarly to cGA. When it stops, each BS makes the final decision on its channel and power level by choosing the locally best sub-chromosome with maximum 'local' fitness, and by setting the channel and power as directed.

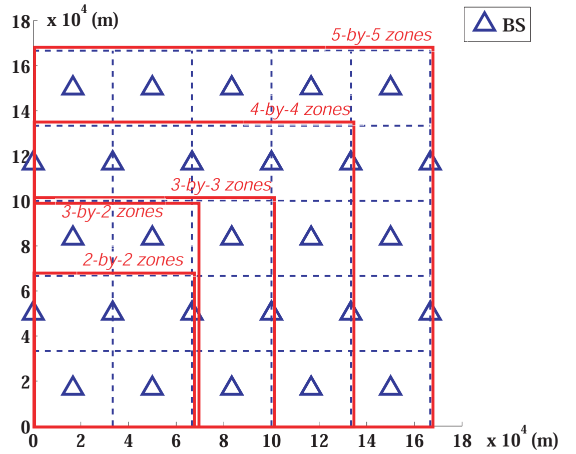


Fig. 7. An illustration of four scenarios: 2×2 to 5×5 zones

C. Implementation Issues on dGA

The communication cost for sub-chromosome exchange is assumed negligible, because the round-trip time of a broadcast packet is just 186 us up to 100 km in 802.22 [26], thus giving 93 μs for one-way communication. Then, it takes only 651 μs for a series of seven broadcast messages.

We propose to use the Coexistence Beacon Protocol (CBP) in 802.22 for broadcast and exchange of the sub-chromosomes between BSs. The CBP is defined in the draft to support *self coexistence* of WRANs, that generates coexistence beacons to perform neighboring network discovery and time synchronization between BSs. Hence, the sub-chromosomes may be piggybacked on CBP packets, by extending the current beacon structure to embed such information.

Resource (re)allocation may occur when network conditions change such as (dis)appearance of a WM/CPE and mobility of a WM. In such cases, up/down-stream communications are temporarily suspended to run the dGA mechanism (for at most 2 seconds), after which BSs may resume their operations by updating the power and channel allocation accordingly.

VII. Performance Evaluation

In this section, our proposed distributed GA (dGA) is evaluated in terms of the achieved fitness at its termination, and compared to the centralized GA (cGA) and the global optimum found by a Brute-Force (BF) search.

A. Simulation Setup

Based on the hexagonal BS deployment model, we consider five scenarios of multi-BS WRANs, as shown in Fig. 7. Each scenario is described by the number of *zones*, each of which is a 33.3 km × 33.3 km rectangular region. For the test scenarios 2×2, 3×2, 3×3, 4×4, and 5×5 zones, the number of BSs (M) is 5, 7, 10, 18, and 27, respectively.

In each zone, L_Z WMs and N_Z CPEs are deployed at random locations. Therefore, $L = L_Z \cdot Z$ and $N = N_Z \cdot Z$

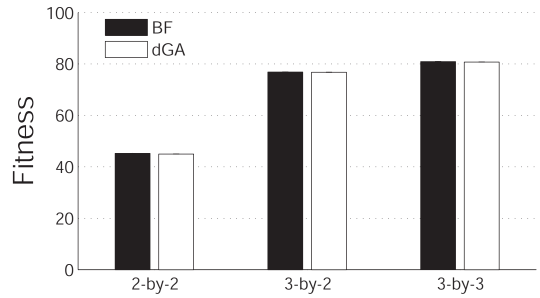


Fig. 8. Performance comparison of dGA with optimal fitness by BF (dGA stops at 2 seconds)

where Z denotes the number of zones. We also assume there are K channels and each WM randomly chooses its channel among them. A test scenario is thus described with (Z, K, N_Z, L_Z), and for the given test we create 10 instances of random WM and CPE deployment. For each deployment, we run cGA and dGA 10 times to study the average behavior.

All simulations are done by using MATLAB running on 2.67 GHz Intel(R) Xeon(R) CPU and 6 Gbytes RAM. Considering that a BS usually has higher computing power than CPEs, the runtime results of dGA presented in this section may become a guideline for the necessary computing power of BSs.

Other simulation parameters are given as follows. For both GA schemes, we use *KEEP_RATE* $= 0.5$, *MUTATION_RATE* $= 0.05$, and $Q = 70$ as the size of a generation. For dGA, we have *EXCHANGE_PERIOD*$= 5$ and *LOCAL_RADIUS*$=$ 40 km. For power allocation, we consider $R_0 \in \mathbf{R} = \{0, 10, 20, \ldots, 70\}$ (in km) thus having $PL^{max} = 8$. For the fan-shaped WM protection regions, i.e., A_2 and A_4, we use $\Theta_1 = \pi/3$ and $\Theta_2 = \pi/6$ in radian. For the ITU-R model, we assume the frequency of 617 MHz (and its adjacent channels), the antenna height of 75 m, and the antenna gain of 12 dBi, as used in [23].

B. Optimal vs. Suboptimal Performance

We compare the performance of dGA measured at 2 seconds of its runtime with the optimal performance found via the brute-force (BF) search, in terms of the maximal fitness achieved. For the comparison, we tested three scenarios, 2×2, 3×2, and 3×3 zones, and used $K = 3$, $N_Z = 3$, and $L_Z = 2$. Note that the computational cost of BF is prohibitive for networks larger than 3×3,[5] limiting the possible test cases. In Fig. 8, dGA is shown to achieve 99.4%, 99.8%, and 99.7% of the optimal fitness for the three scenarios tested, respectively. That is, the dGA efficiently achieved near-optimal performance.

C. Centralized vs. Distributed GA

We now compare the performance of cGA with dGA, both stopping at 2 seconds of runtime. For this comparison, we tested the four scenarios, 2×2, 3×3, 4×4, and 5×5 zones, and used $K = 3$, $N_Z = 3$, and $L_Z = 2$. As shown in Fig. 9,

[5]The runtimes for the three tested cases are 6, 231, and 9,282 seconds, which grows as fast as K^M.

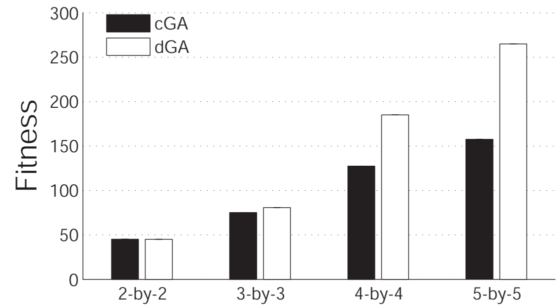


Fig. 9. Performance comparison of dGA with cGA (both stop at 2 seconds)

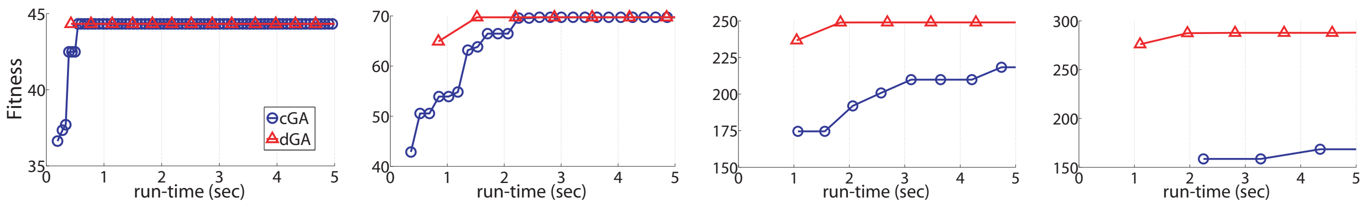
cGA performs comparable to dGA for small networks such as 2×2 and 3×3. As the network size grows, however, dGA significantly outperforms cGA as seen from 4×4 and 5×5. This is because cGA has to deal with a larger chromosome involving channel and power allocation of all BSs, resulting in a slower evolution.

The speed of fitness evolution is clearly shown in Fig. 10, where we let cGA and dGA run for more than 2 seconds and plot the instantaneous maximal fitness of the two schemes as a function of runtime. For better readability, dGA plots the fitness once every *EXCHANGE_PERIOD* iterations (*EXCHANGE_PERIOD*$= 5$), and cGA plots its fitness for each iteration. cGA is found to take longer to converge as the network size grows while dGA can converge within 2 seconds for all cases. The performance gap between two at 2 seconds becomes more pronounced as the network size M grows, especially for 4×4 and 5×5, making 3×3 a triggering condition beyond which dGA is favored. In addition, as M increases, dGA shows only a moderate increase in the completion time of 5 iterations, whereas cGA shows a dramatic increase in its inter-iteration time. Note that one marker in dGA corresponds to five markers in cGA, and thus dGA completes each iteration much faster than cGA.

D. Distributed GA for Various Scenarios

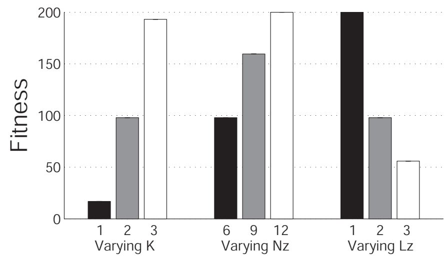
We also measured dGA's performance at 2 seconds under various test conditions and compared it with the optimal fitness by BF. For this, we considered 3×3 zones and varied one of K, N_Z, and L_Z while fixing others, resulting in the following three scenarios: (1) $K = 1, 2, 3$, $N_Z = 6$, $L_Z = 2$, (2) $K = 2$, $N_Z = 6, 9, 12$, $L_Z = 2$, and (3) $K = 2$, $N_Z = 6$, $L_Z = 1, 2, 3$. Considering the fact that IEEE 802.22 allows a total of 12 simultaneous CPEs per channel [10], the chosen scenarios represent the reality of the problem very well.

Fig. 11(a) shows that the achieved fitness gradually increases as K grows due to the channel diversity, where each BS has more freedom in deciding its channel while avoiding conflicts with existing WMs and other WRANs. As a result, BSs can use higher power providing higher SNR at CPEs. Next, the achieved fitness increases almost linearly as N_Z grows, because the fitness function is the sum-rate of all CPEs with $N = N_Z \cdot Z$. Finally, the achieved fitness decreases as L_Z grows, since the blackout regions by A_2 and A_4 increases and thus BSs should reduce their power and CPEs get harder to find a BS to associate with.

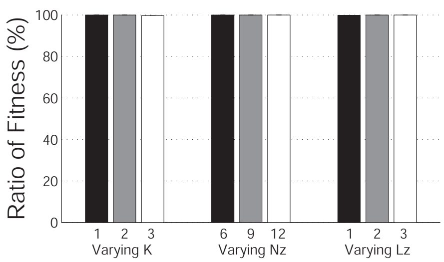


(a) 2×2 case (b) 3×3 case (c) 4×4 case (d) 5×5 case

Fig. 10. Iterative evolution and convergence of fitness: cGA vs. dGA



(a) Fitness of dGA



(b) Ratio of Fitness between dGA and BF

Fig. 11. Performance of dGA under various conditions

Fig. 11(b) illustrates the ratio (in %) of the achieved fitness by dGA compared to the BF's optimal fitness. The ratio gives more than 99.85% for tests (2) and (3), while it gradually decreases as K grows at test (1) from 100% to 99.6%. It shows that for a given M, the major overhead of dGA comes from the number of channels (K), not from N or L.

VIII. Conclusion

In this paper, we formulated the problem of joint allocation of channel, power, and CPEs in IEEE 802.22 as a MINLP. To solve this NP-hard problem in real time to conform to the FCC regulations, i.e., at most 2 seconds for spectrum re-allocation, we proposed and applied a distributed Genetic Algorithm (dGA) that is computationally efficient and provides reasonably good performance as compared to both the optimal solution and the centralized GA. In future, we would like to extend this approach to combine small- and large-scale PUs in the same framework to investigate its impact on the performance of joint resource allocation. We are also interested in exploring the performance of other meta-heuristic algorithms, such as tabu search, simulated annealing, ant colony optimization, and particle swarm optimization.

Acknowledgments

The work described in this paper was supported in part by the NSF under Grant CNS-0721529.

References

[1] FCC. Facilitating Opportunities for Flexible, Efficient, and Reliable Spectrum Use Employing Cognitive Radio Technologies. *ET Docket No. 03-108*, December 2003.
[2] FCC. Notice of Proposed Rulemaking and Order. *ET Docket No. 04-113*, May 2004.
[3] S. Shellhammer et al. Performance of power detector sensors of DTV signals in IEEE 802.22 WRANs. In *ACM TAPAS*, Aug. 2006.
[4] F.F. Digham. Joint Power and Channel Allocation for Cognitive Radios. In *IEEE WCNC 2008*, pages 882–887, April 2008.
[5] Y. Che et al. A Two-Step Channel and Power Allocation Scheme in Centralized Cognitive Networks Based on Fairness. In *IEEE VTC Spring*, pages 1589–1593, May 2008.
[6] H. He et al. Game Theoretic Analysis of Joint Channel Selection and Power Allocation in Cognitive Radio Networks. In *Proceedings of the CrownCom 2008*, May 2008.
[7] J. Li et al. Multiuser Power and Channel Allocation Algorithm in Cognitive Radio. In *IEEE ICPP*, Sep. 2007.
[8] M. Chatterjee, H. Lin, and S.K. Das. Rate Allocation and Admission Control for Differentiated Services in CDMA Data Networks. *IEEE T-MC*, 6(2):179–191, February 2007.
[9] M.Y. ElNainay et al. Channel Allocation & Power Control for Dynamic Spectrum Cognitive Networks using a Localized Island Genetic Algorithm. In *IEEE DySPAN*, pages 1–5, October 2008.
[10] C. Cordeiro et al. IEEE 802.22: An Introduction to the First Wireless Standard based on Cognitive Radios. *Journal of Communications*, 1(1):38–47, April 2006.
[11] E. Sofer and G. Chouinard. WRAN Channel Modeling. *IEEE 802.22-05/0055r7*, Sep. 2005.
[12] S. Shellhammer et al. Spectrum Sensing Simulation Model. *IEEE 802.22-06/0028r10*, September 2006.
[13] W. Caldwell. Draft Recommended Practice. *IEEE 802.22-06/0242r6*, March 2007.
[14] BeamFlex, http://www.ruckuswireless.com.
[15] M. Brown. Part 22.1: Enhanced Protection for Low-Power, Licensed Devices Operating in Television Broadcast Bands. *IEEE 802.22-06/0238r2*, March 2007.
[16] W. Caldwell. Geolocation with Database Requirement Summary. *IEEE 802.22-06/0140r0*, July 2006.
[17] H.-S. Chen et al. Spectrum Sensing for Wireless Microphone Signals. In *IEEE SECON 2008*, June 2008.
[18] H. Kim and K.G. Shin. In-band Spectrum Sensing in Cognitive Radio Networks: Energy Detection or Feature Detection? In *Proceedings of the ACM MobiCom 2008*, September 2008.
[19] G. Chouinard. Wireless Microphone Sensing. *IEEE 802.22-07/0530r1*, November 2007.
[20] C.-J. Kim et al. Fractional BW Usage for WRAN Systems. *IEEE 802.22-06/0117r0*, July 2006.
[21] E. Au and G. Chouinard. Draft Minutes of the Orlando Plenary Session of 802.22. *IEEE 802.22-08/0112r0*, March 2008.
[22] ITU-R Recommendation P.1546-3. Method for point-to-area predictions for terrestrial services in the frequency range 30 MHz to 3000 MHz.
[23] G. Chouinard. WRAN Reference Model. *IEEE 802.22-04/0002r12*, September 2005.
[24] M.D. Vose. *The Simple Genetic Algorithm: Foundations and Theory*. The MIT Press, Cambridge, MA, 1999.
[25] R. Leardi. *Nature-inspired Methods in Chemometrics: Genetic Algorithms and Artificial Neural Networks*. Elsevier Science, Inc., New York, NY, 2003.
[26] C.R. Stevenson et al. Functional Requirements for the 802.22 WRAN Standard. *IEEE 802.22-05/0007r20*, May 2005.