

A Framework for Cognitive WiMAX with Frequency Agility

Alexe E. Leu *Member, IEEE*, Brian L. Mark *Member, IEEE*, and Mark A. McHenry *Senior Member, IEEE*

Abstract—Cognitive radios have the ability to sense the radio spectrum environment and to switch dynamically to available frequency ranges. Mobile WiMAX is an emerging wireless networking standard which could potentially benefit from cognitive radio technology. We develop a framework for applying cognitive radio technology to mobile WiMAX networks to increase capacity and simplify network operations. In the proposed cognitive WiMAX architecture, base stations are equipped with sensitive detectors and assign channels to subscriber stations dynamically based on spectrum availability. Power control is employed to increase frequency reuse in conjunction with spectrum sensing. Using computer simulation, we evaluate the performance of “cognitive channel assignment” relative to conventional dynamic channel assignment. Our numerical results show that cognitive radios can substantially increase the capacity of emerging WiMAX networks by exploiting inherent spectrum hole opportunities. The key performance parameters determining the achievable capacity of cognitive WiMAX networks are the detection and interference range, which depend in turn on characteristics of the radio propagation environment.

Index Terms—WiMAX, IEEE 802.16, cognitive radio, frequency-agile radio, spectrum sensing, power control.

I. INTRODUCTION

During the past decade, broadband wireline and wireless mobile services have been two of the most remarkable growth areas in the telecommunications industry. Broadband access has been enabled by technologies such as Digital subscriber line (DSL), cable modems, and fiber-to-the-home (FTTH). Wireless mobile access has been driven by 2G and now 3G cellular systems. In addition to cellular systems, Wi-Fi systems based on the IEEE 802.11 family of standards, have become enormously popular for providing in-building wireless coverage. Wi-Fi systems offer much higher peak data rates than 3G systems, but are not designed to support high-speed mobility.

WiMAX is an emerging technology based on the IEEE 802.16 family of standards for broadband wireless mobile access that provides a rich set of features and a high degree of flexibility [1], [2]. In this paper, the term “WiMAX” refers to “mobile WiMAX,” i.e., the WiMAX standard that accommodates mobile subscribers. Unlike 3G systems, which provide only a fixed channel bandwidth, WiMAX allows the

user to select an adjustable channel bandwidth from 1.25 MHz to 20 MHz. By using Orthogonal Frequency Division Multiplexing (OFDM) as the primary modulation scheme, WiMAX, as well as Wi-Fi, is able to support much higher peak data rates than 3G systems that are based on Code Division Multiple Access (CDMA), which requires bandwidth spreading. OFDM is a multicarrier modulation scheme whereby a given high-rate data stream is divided into several parallel low bit-rate streams, each of which is modulated onto a separate carrier called a subcarrier or tone. The multiple access scheme adopted by WiMAX is Orthogonal Frequency Division Multiple Access (OFDMA), whereby the available subcarriers are further divided into groups called subchannels, which can be allocated to different users. In OFDMA, the subcarriers assigned to a subchannel need not be contiguous, allowing for a flexible assignment of data rates to users.

In recent years, dynamic spectrum access (DSA) has been an active area of research [3]–[7] because of its potential to exploit highly underutilized wireless spectrum. Cognitive radios with frequency agility enable DSA by sensing spectrum “holes” and automatically tuning to available frequency channels. Much of the research on DSA has focused on systems consisting of secondary users equipped with cognitive radios that attempt to utilize spectrum that is not being used by the primary licensed users at a particular time and place. The key technological requirements for frequency-agile cognitive radios include (i) highly sensitive detectors to accurately detect the presence or absence of transmissions from primary users, (ii) mechanisms to determine the maximum interference-free transmit power (MIFTP) [3], and (iii) methods to negotiate common frequency channels with other secondary users and to quickly tune to selected channels.

In this paper, we investigate the application of frequency-agile cognitive radios to WiMAX systems to increase network capacity and to simplify the network infrastructure. We refer to a cognitive radio-enabled WiMAX system as *cognitive WiMAX*. In a cognitive WiMAX system, base stations are equipped with cognitive radios that sense the spectrum availability and dynamically assign channel resources among the CR nodes. We propose an architecture for cognitive WiMAX based on a *cognitive channel assignment (CCA)* scheme and study the achievable capacity improvements over conventional WiMAX system architectures.

The remainder of the paper is organized as follows. Section II provides an overview of the relevant aspects of WiMAX and cellular systems. Section III discusses the key frequency-agile cognitive radio technologies and mechanisms that enable cognitive WiMAX. Section IV proposes an architecture for

Manuscript received January 31, 2008.

A. E. Leu and M. A. McHenry are with Shared Spectrum Company, Vienna, VA 22182. (e-mail: aleu@sharedspectrum.com, mmchenry@sharedspectrum.com).

B. L. Mark is with the Dept. of Electrical and Computer Engineering, The Volgenau School of Information Technology & Engineering, George Mason University, Fairfax, VA 22030 (e-mail: bmark@gmu.edu).

This work was supported in part by the U.S. National Science Foundation under Grant CNS-0520151.

cognitive WiMAX systems based on frequency-agile CRs employing CCA. The capacity improvement of cognitive WiMAX over conventional WiMAX is investigated quantitatively via computer simulation in Section V. Finally, the paper is concluded in Section VI.

II. BACKGROUND ON WiMAX AND CELLULAR SYSTEMS

In this section, we provide a brief overview of WiMAX and relevant concepts in cellular systems. For further details on WiMAX, which stands for the Worldwide Interoperability for Microwave Access, reader is referred to white papers from the WiMAX forum [8], [9], overview papers [10], and recent books on WiMAX [1], [2]. Cellular systems are discussed in many books on wireless networks [11], [12].

The WiMAX Forum is promoting broadband wireless technology based on the IEEE 802.16 family of standards. The original 802.16 standard for fixed wireless access was completed in 2001 based on a single-carrier physical (PHY) layer with a medium access control (MAC) layer based on TDMA (time division multiple access). Subsequently, 802.16a was developed, based on *orthogonal frequency division multiplexing* (OFDM) and *orthogonal frequency division multiple access* (OFDMA). Further revisions resulted in a new standard released in 2004 called IEEE 802.16-2004 [13], which replaced all prior versions of 802.16. This standard formed the basis of the first WiMAX standard, referred to as *fixed WiMAX*. In 2005, a new standard called 802.16e-2005 [14] was completed, which included mobility support. This standard forms the basis for *mobile WiMAX* technology. In the remainder of this paper, we shall assume mobile WiMAX, but refer to both fixed and mobile WiMAX collectively as simply WiMAX.

A. WiMAX Physical Layer

The WiMAX PHY specifies several operational frequency bands, including 2–11 GHz for fixed applications [13] and 2–6 GHz for mobile applications [14]. An earlier version of the 802.16 standard specified a frequency band of 10–60 GHz for fixed applications [15]. The PHY layer is based on OFDM, a multicarrier modulation scheme which provides strong mitigation of multipath effects and allows for operation in non-line-of-sight (NLOS) conditions. In OFDM, a high bit-rate stream is divided into several parallel low bit-rate substreams, each of which is modulated onto a separate subcarrier or tone. The substream symbol time is chosen to be large enough so that the delay spread incurred by the wireless channel is a small fraction of the symbol duration, thus minimizing intersymbol interference (ISI). The subcarriers are chosen to be mutually orthogonal over the symbol period, such that the subcarrier channels need not be nonoverlapping.

An OFDM signal can be generated by taking the inverse Discrete Fourier Transform (IDFT) of the input data stream in blocks of L symbols, where L is the number of subcarriers. OFDM transmitters and receivers can be implemented with low complexity using the Fast Fourier Transform (FFT). Besides ISI-mitigation and low computational complexity, OFDM provides frequency diversity by allowing

coding/interleaving across subcarriers and robustness against narrowband interference.

The multiple access technique used in WiMAX is called scalable OFDMA because the FFT size used in OFDM can be scaled from 128 to 2048 [13], [14]. As the available spectrum bandwidth increases, the FFT size for OFDM can be increased to maintain a constant subcarrier spacing. Typically, the subcarrier spacing is 10.94 kHz. Thus, when the channel bandwidth is 1.25, 5, 10, and 20 MHz, the FFT size is set to 128, 512, 1024, and 2048, respectively, to maintain the 10.94 kHz subcarrier spacing.

In OFDMA, the available subcarriers are further partitioned into groups of subcarriers called *subchannels*. Different subchannels are assigned to different users. By assigning different numbers of subcarriers to subchannels, fine-grained resource allocation can be achieved. The subcarriers making up a subchannel need not be contiguous. Subchannels consisting of noncontiguous subcarriers offer greater frequency diversity. WiMAX defines several different subchannelization schemes for both the uplink and downlink.

OFDMA may be considered a hybrid of TDMA (Time Division Multiple Access) and FDMA (Frequency Division Multiple Access) (cf. [2]), in the sense that users are allocated both OFDM subcarriers and time slots. OFDMA offers the multipath suppression and frequency diversity of OFDM plus flexible allocation of rates to users (cf. [16]). In the time domain, data is transmitted in the form of frames. The minimum time-frequency unit that can be allocated to a user is a *slot*. A slot consists of a subchannel over one, two, or three OFDM symbols, depending on the subchannelization scheme that is used. A contiguous set of slots assigned to a user is called a *data region*.

WiMAX supports both TDD (Time Division Duplexing) and FDD (Frequency Division Duplexing). Fig. 1 shows sample WiMAX OFDMA frame structures for FDD and TDD. Fig. 1(a) shows a sample WiMAX TDD frame structure (cf. [2]) consisting of 6 data regions on the downlink and uplink, respectively. Similarly, Fig. 1(b) shows a WiMAX FDD frame structure with 6 data regions. In both cases, the uplink and downlink media access protocol (MAP) messages (UL-MAP and DL-MAP) specify the allocation of users to data regions within the frame. The ranging channel in the uplink portion of the WiMAX frame provides contention-based access for frequency, time, and power adjustments. The ranging channels can also be used by a mobile station (MS) to make uplink bandwidth requests. Best-effort traffic may also be transmitted on the ranging channel when the amount of data to be sent is relatively small. The frame size can be varied on a frame-by-frame basis from 2 to 20 ms, but the nominal frame size is 5 ms. With an OFDM symbol duration of 102.9 μ s, the number of OFDM symbols in a 5 ms frame is 48. In general, TDD is the preferred duplexing method because it allows for simpler and more flexible sharing of bandwidth between uplink and bandwidth (cf. [2]). On the other hand, TDD requires synchronization across multiple base stations.

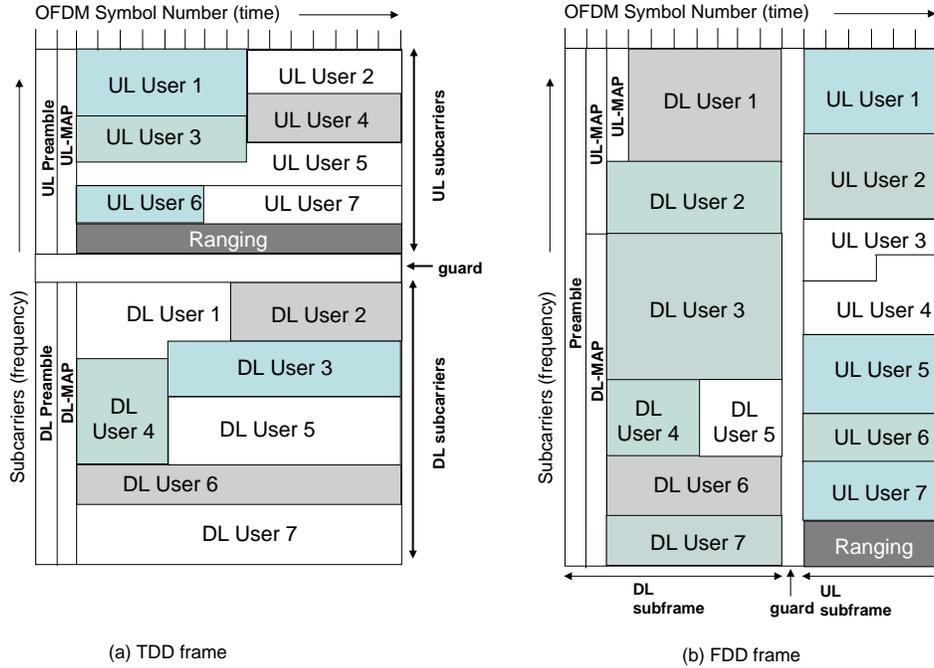


Fig. 1. Sample WiMAX frame structures. (a) TDD; (b) FDD.

B. WiMAX MAC Layer

The WiMAX MAC layer takes packets from the upper layer, called MAC service data units (MSDUs) and transforms them into MAC protocol data units (MPDUs) for transmission over the PHY layer. The MAC layer includes a convergence sublayer that can provide an interface to various high layer protocols, but currently only IP and Ethernet are supported. In WiMAX, the base station is responsible for allocating bandwidth to all users on the uplink and downlink. On the downlink, the BS allocates bandwidth to each MS according to the requirements of the incoming traffic without involving the MS. On the uplink, the MS makes requests for uplink bandwidth via a polling mechanism overseen by the BS.

The WiMAX MAC layer is connection-oriented in the sense that prior to data transmission, a logical link called a *connection* must be established between the BS and the MS. The connection is assigned a connection identifier (CID). The connection-oriented architecture allows WiMAX to support fine-grained Quality-of-Service (QoS). A *service flow* is a unidirectional flow of packets associated with a set of QoS parameters and identified by a *service flow identifier* (SFID). The QoS parameters include priority, maximum sustained traffic rate, minimum tolerable rate, maximum delay, etc.

WiMAX specifies five scheduling services summarized below [2], [17]:

- 1) **Unsolicited grant services (UGS):** This service supports constant bit rate (CBR) traffic with fixed-size data packets.
- 2) **Real-time polling services (rtPS):** This service supports real-time variable bit rate (VBR) traffic flows that generate variable-size data packets on a periodic basis.

- 3) **Non-real-time polling services (nrtPS):** This service supports delay-tolerant flows that require a minimum guaranteed traffic rate.
- 4) **Best-effort service (BE):** This service supports data streams that do not require minimum QoS guarantees.
- 5) **Extended real-time variable rate (ERT-VR) service:** This service supports real-time traffic flows that require a guaranteed data rate and delay.

While WiMAX provides extensive bandwidth allocation and QoS mechanisms, the details of scheduling and management are left unstandardized and provide an important avenue for vendors to differentiate their equipment.

C. Cellular System Concepts

WiMAX systems are expected to be deployed primarily as cellular systems wherein the geographic coverage area is partitioned into smaller regions called *cells*. Each cell is served by a base station, which limits its transmit power to provide sufficient signal strength at the cell boundary. Propagation path loss allows base stations in spatially separated cells to transmit at the same carrier frequencies without causing harmful interference to each other.

In conventional cellular systems based on FDMA (Frequency Division Multiple Access), the system bandwidth is divided into frequency channels of equal bandwidth. Each channel provides a communication link for a single connection or call. If frequency-division duplexing (FDD) is used, separate frequency channels must be allocated for the uplink and downlink channels. In time-division duplexing (TDD), a single frequency channel supports both the uplink and downlink channels via time-division multiplexing.

As discussed above, WiMAX is based on OFDMA, which allows a very flexible allocation of spectrum to users to accommodate different traffic types and data rate requirements. In OFDMA, subcarriers are grouped into subchannels which are allocated to users. From the user's perspective, an OFDMA subchannel corresponds to a frequency channel in conventional FDMA-based cellular systems, except that the bandwidth of an OFDMA subchannel can be variable. To avoid co-channel interference in OFDMA, however, the *subcarrier* would be the basic unit of frequency allocation.

To simplify the discussion, we shall discuss frequency allocation in cellular systems in terms of frequency channels in conventional FDMA-based cellular systems with the understanding that for WiMAX, frequency allocation would be performed at the granularity of a subcarrier. The performance study presented in Section V is given in terms of frequency channels for the sake of simplicity and to facilitate comparison with conventional cellular systems. The cognitive radio techniques discussed in this paper are applicable to WiMAX as well as FDMA-based cellular systems. Ultimately, however, the payoff in applying cognitive radio technology is much greater for WiMAX than for FDMA systems due to the finer granularity of spectrum allocation in WiMAX.

The mechanism used to assign frequency channels within a cell in an FDMA cellular system is referred to as a *channel assignment scheme*. Two channel assignment schemes that have been used in conventional cellular networks are *fixed channel assignment* (FCA) and *dynamic channel assignment* (DCA). In Section IV, we shall introduce a new channel assignment scheme based on cognitive radio technology, which we shall refer to as *cognitive channel assignment* (CCA).

1) *Fixed Channel Assignment (FCA)*: In FCA [11], [12], [18]–[21], the coverage area is partitioned into groups of contiguous cells called *clusters*. The set of frequency channels is partitioned evenly among the cells in any given cluster such that each cell in the network is allocated a predetermined set of channels. Any call request within the cell can only be served by the unused channels assigned to that particular cell.

To improve utilization, a borrowing option may be considered [22]. With the borrowing option, a cell is allowed to borrow channels from a neighboring cell if all of its own channels are already occupied and the neighboring cell has spare channels. Borrowing is normally supervised by the mobile switching center (MSC). Since handoff is performed by the MSC, the MSC has full knowledge of the capacity usage of the cluster of cells within its jurisdiction. Therefore, the MSC is the natural subsystem to oversee functions such as channel borrowing.

2) *Dynamic Channel Assignment (DCA)*: In DCA [23]–[30], channels are not allocated to cells on a permanent basis. Each time a call request is made, the serving base station requests a channel from the MSC. The MSC dynamically determines the availability of a channel and executes its allocation procedure accordingly. The MSC only allocates a given frequency channel if that channel is not presently in use in the cell, or any other cell which falls within the minimum restricted distance of frequency reuse to avoid co-channel interference.

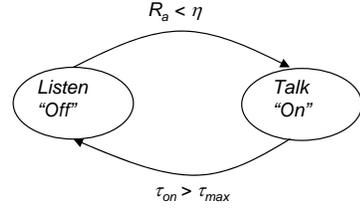


Fig. 2. State transition diagram for LBT algorithm.

The DCA scheme can be explained in terms of the cell cluster concept (cf. [11]). For a given cell i , an associated MSC maintains a list of channels with an indication of whether the channel is free or occupied. When a call request arrives to cell i , the call is assigned a free channel, say channel c , if one is available. In this case, channel c is marked as *occupied* for all of the other cells in the cell cluster centered at cell i . Later, when the call completes, channel c is marked as *free* for all cells in the cluster centered at cell i .

DCA reduces the likelihood of call blocking, which increases the trunking capacity of the system, since all available channels under control of the MSC are accessible to all of the cells. On the other hand, DCA schemes require the MSC to collect real-time data on channel occupancy, traffic distribution, and radio signal quality of all channels on a continuous basis. The MSC needs to do this data collection in order to manage handoffs between cells.

III. FREQUENCY-AGILE COGNITIVE RADIO TECHNOLOGY

In this section, we review some basic concepts of frequency-agile cognitive radio technology, which will be used in the development of the cognitive WiMAX architecture proposed in Section IV. A *frequency-agile* cognitive radio (CR) is capable of sensing the spectrum and dynamically tuning to frequency channels determined to be available [31]. In [3], a framework was developed to evaluate the performance of dynamic spectrum access (DSA) schemes, and in particular, the Listen-Before-Talk (LBT) scheme. For simplicity, in this paper, we shall assume that the CR nodes employ *Listen-Before-Talk* (LBT) dynamic spectrum access. As discussed in [3], more substantial spectrum gains can be achieved by schemes in which the CR nodes collaborate with each other to perform dynamic spectrum access, at the expense of higher communication and computational complexity.

A. Listen-Before-Talk Spectrum Access

The Listen-Before-Talk (LBT) algorithm is a scheme for a CR node to access a radio frequency channel dynamically. The LBT scheme consists of two states: (1) listen or *off* state; and (2) the talk or *on* state. During the off state, CR node does not transmit a signal and estimates the received signal power R in the radio channel c . In the off state, the CR node also estimates a transmit power level, s^* , which we refer to as the maximum interference-free transmit power (MIFTP) [3].

The MIFTP is defined as the maximum power at which the CR node can transmit without causing harmful interference to

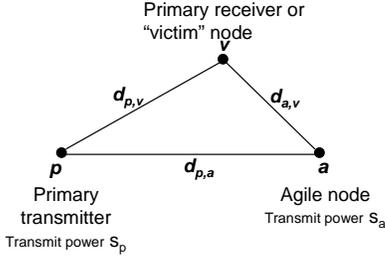


Fig. 3. Interference scenario.

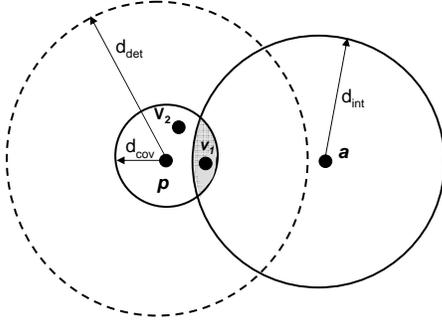


Fig. 4. Interference condition.

any of the primary nodes. We also refer to the primary nodes as victim or non-cooperative nodes. During the off state, the CR node listens to the channel and transitions to the on state if $R < \eta$. Otherwise, the CR node remains in the off state for the same channel c or switches to a different frequency channel c' that may be available. For simplicity, we shall consider only the case where the CR node seeks to use the same channel c . During the on state, the CR node transmits at a power level less than or equal to the MIFTP s^* for a maximum duration of t_{\max} , and then returns to the off state to listen again. Fig. 2 illustrates the LBT algorithm by means of a state transition diagram. More sophisticated variations of the LBT scheme can be devised that involve collaboration among a group of CR nodes [3], [32].

Consider a scenario consisting of three nodes shown in Fig. 3: a primary transmitter p , a victim receiver v , and a CR node denoted by a . Node p transmits on frequency channel c , while node a performs LBT on the channel. The receiver v expects to receive the transmissions of node p and is a potential *victim* node of the CR node a since node a may cause interference to node v while in the on state.

We introduce several quantities of interest with respect to the simple LBT scenario of Fig. 3 (cf. [3]). Assume that the primary node p transmits with constant power s_p . The received power at node v from node p is given by

$$R_v \triangleq s_p - L_{p,v}, \quad (1)$$

where $L_{p,v}$ denotes the propagation loss from node p to node v . The propagation loss between two nodes i and j can be modeled as a sum of two components:

$$L_{i,j} = D_{i,j} + W_{i,j}, \quad (2)$$

where $D_{i,j}$ denotes the median path loss and $W_{i,j}$ is a random variable representing shadowing noise. The path loss component $D_{i,j}$ depends on the terrain profile between nodes i and j . Examples of path loss models that are relevant to WiMAX deployments [2] include the Okumura-Hata model, the COST-231 Hata model, the Walfish-Ikegami model, and the Erceg model. These models are suitable for both Line-of-Sight (LOS) and Non-Line-of-Sight (NLOS) scenarios and are based on empirical measurements. The simulation results discussed in Section V assume an Erceg path loss model [33].

The EPM-73 propagation model [34] is a representative empirical path loss model in which the path loss depends only on the terrain *type* and the distance between the nodes, rather than on the detailed terrain profile. Assuming a path loss model similar to EPM-73, the path loss can be expressed as a function of the distance $d_{i,j}$ between nodes i and j , the height $h_i^{(t)}$ of the transmit antenna at node i , the height $h_j^{(r)}$ of the receive transmitter at node j , the carrier frequency f_c , and the antenna polarization o_i at node i :

$$D_{i,j} = g(d_{i,j}, h_i^{(t)}, h_j^{(r)}, f_c, o_i). \quad (3)$$

To simplify the notation, we shall write $D_{i,j} = g(d_{i,j})$, suppressing the dependence of the path loss on the remaining four parameters. Assuming the function $g(\cdot)$ is invertible, the distance between nodes i and j can be obtained from the path loss $D_{i,j}$ as follows:

$$\tilde{d}_{i,j} = g^{-1}(D_{i,j}). \quad (4)$$

The shadowing noise $W_{i,j}$ is typically modeled as a zero-mean white Gaussian noise process with variance $\sigma_{i,j}^2$, which is independent of the path loss $D_{i,j}$.

The *outage probability* of node v with respect to the primary node p is the probability that the received signal strength R_v falls below a threshold r_{\min} :

$$P_{\text{out}} \triangleq P\{R_v < r_{\min} | E_p\}, \quad (5)$$

where E_p denotes the event that node p is in the on state. The *coverage distance* of node p is the maximum distance between the node p and any victim node v such that the outage probability does not exceed a value ϵ_{out} :

$$d_{\text{cov},p}(\epsilon_{\text{out}}) \triangleq \max\{d_{p,v} : P_{\text{out}} \leq \epsilon_{\text{out}}\}. \quad (6)$$

The outage probability and coverage distance depend on the propagation loss between the primary node p and the victim node v . With the assumptions on propagation loss discussed above, the coverage distance can be expressed as [3]:

$$d_{\text{cov},p}(\epsilon_{\text{out}}) = g^{-1}(s_p - r_{\min} + \alpha_{p,v}), \quad (7)$$

where

$$\alpha_{p,v} \triangleq \sigma_{p,v} Q^{-1}(1 - \epsilon_{\text{out}}), \quad (8)$$

where $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$ denotes the standard Q-function (cf. [11]). The *coverage region* associated with the primary node p is the area enclosed by the circle centered at node p with radius $d_{\text{cov},p}$.

Let R_a denote the received power at the CR node a from the primary node p . We define the detection probability, P_{det} ,

at the CR node as the probability that the received signal from the primary node p is greater than or equal to a given threshold η_{det} :

$$P_{\text{det}} \triangleq P\{R_a \geq \eta_{\text{det}}\}. \quad (9)$$

The detection distance, d_{det} , at the CR node a is defined as the maximum distance between the CR node and the primary node such that the probability of detection exceeds a value ϵ_{det} :

$$d_{\text{det}}(\epsilon_{\text{det}}) \triangleq \max\{d_{a,p} : P_{\text{det}} \geq \epsilon_{\text{det}}\}. \quad (10)$$

The detection distance can be expressed as follows:

$$d_{\text{det}}(\epsilon_{\text{det}}) = g^{-1}(s_p - \eta_{\text{det}} - \alpha_{p,a}), \quad (11)$$

where

$$\alpha_{p,a} \triangleq \sigma_{p,a} Q^{-1}(1 - \epsilon_{\text{det}}). \quad (12)$$

The *detection region* is the area enclosed by the circle centered at node a with radius d_{det} .

Let I_v denote the interference power received at the victim node v from the CR node a :

$$I_v \triangleq s_a - L_{a,v}. \quad (13)$$

Harmful interference occurs when the signal power received at node v from node a exceeds a threshold i_{max} and the received power R_v exceeds r_{min} . Thus, the *interference probability*, P_{int} , is given by

$$P_{\text{int}} = P\{R_v \geq r_{\text{min}}, E_p\} \cdot P\{I_v > i_{\text{max}}, E_a\}, \quad (14)$$

where E_a denotes the event that node a is in the on state and we assume independence of R_v and I_v .

Let ϵ_{int} denote the maximum interference probability that can be tolerated by the victim node v . The interference distance is defined as the minimum permissible distance between the CR node a and the victim node v such that the interference probability does not exceed ϵ_{int} :

$$d_{\text{int}}(\epsilon_{\text{int}}) \triangleq \min\{d_{a,v} : P_{\text{int}} \leq \epsilon_{\text{int}}\}. \quad (15)$$

We define the *interference region* of the CR node a to be the region enclosed by the circle centered at node a with radius d_{int} . The interference distance can be expressed as follows [3]:

$$d_{\text{int}} = g^{-1}(s_a - i_{\text{max}} - \alpha_{a,v}), \quad (16)$$

where

$$\alpha_{a,v} \triangleq \sigma_{a,v} Q^{-1} \left(1 - \frac{\epsilon_{\text{int}}}{(1 - P_{\text{out}})(1 - P_{\text{det}})} \right). \quad (17)$$

The coverage, detection, and interference distances are key parameters in determining the performance of the LBT spectrum access scheme. In [3], it is shown that if the outage probability and detection probabilities satisfy

$$P_{\text{out}} \leq \epsilon_{\text{out}} \text{ and } P_{\text{det}} \leq \epsilon_{\text{det}}, \quad (18)$$

then the victim node does not suffer harmful interference if the interference distance is less than or equal to the detection distance minus the coverage distance:

$$d_{\text{int}} \leq d_{\text{det}} - d_{\text{cov},p}. \quad (19)$$

We refer to (19) as the *noninterference condition*.

In Fig. 4, the CR node a lies just outside the detection distance d_{det} of the primary node p and the two victim nodes v_1 and v_2 lie within the coverage region of node p . Observe that node v_2 would not suffer harmful interference, since it lies beyond the interference region of the CR node a . However, the victim node v_1 lies within the coverage region of node p and within the interference region of node a . Hence, node v_1 would potentially suffer harmful interference from node a . Furthermore, any victim node lying within the shaded region would suffer harmful interference. The shaded region in Fig. 4 can be reduced by lowering the CR node transmit power s_a . If the coverage region of node p and the interference region of node a do not intersect, i.e., the shaded region is absent, then the noninterference condition (19) is satisfied for all potential victim nodes.

B. Signal Detectors

As can be inferred from (19), one way of avoiding a co-channel interference condition is to increase the value of d_{det} . From (10), we see that d_{det} can be increased by reducing the detection threshold η_{det} . Thus, ultra-sensitive signal detectors play an important role in frequency agile radio technology (cf. [31], [35], [36]).

To achieve high utilization in cognitive WiMAX, the processing time of the detector should be as small as possible. Assuming a WiMAX TDD frame duration of 5 ms, the detector would need one frame time to determine the availability of a given subchannel. Wideband detectors can achieve a processing time of less than 2 ms, but are also the most expensive. Slower detectors could be used by amortizing the cost of sensing over one frame time over the lifetime of a connection assigned to a given subchannel; the longer the connection lifetime, the smaller the bandwidth wastage. Thus, the slower detectors can be efficient in terms of bandwidth utilization provided that the connection holding time is significantly larger than the detector processing time.

IV. COGNITIVE WiMAX

In this section, we propose a framework, called *cognitive WiMAX*, for applying frequency-agile cognitive radio technology to WiMAX networks to increase frequency reuse and network capacity, and to simplify network operations. The cognitive radio framework developed here can also be applied to conventional FDMA-based cellular networks. However, the impact of applying cognitive radio technology will be greater for WiMAX due to its support for fine-grained and flexible frequency allocation. In cognitive WiMAX, each base station is equipped with a frequency-agile CR and can assume the role of a primary node, a CR node, or a victim node with respect to dynamic spectrum access.

A. Related Work

A broad survey of fixed, dynamic, and hybrid allocation schemes, and discussions of other issues such as overlay cells, frequency planning, and power control is provided in [22]. In

[37], a framework for the design and evaluation of frequency-agile MAC protocols is proposed. In [38], a MAC protocol based on partially observable Markov decision processes is proposed for ad hoc networks with opportunistic spectrum access.

System capacity performance issues are studied in [39] through the signal-to-interference ratio in both the uplink and downlink. Zander [40] develops a distributed dynamic channel assignment optimum algorithm using signal-to-interference ratio measurements such that capacity gains on the order of 3-4 are feasible for systems with slowly varying link gains. In [41], Zander provides capacity bounds showing that dynamic channel assignment (DCA) algorithms in the asymptotic case is just above twice the capacity of a fixed channel assignment (FCA) scheme. The optimum capacity can be achieved using various approaches such as neural networks [42], [43], simulating annealing [44], and genetic algorithms [45]. Results presented in [46]–[50] show that the best DCA strategies are those based on the interference levels measured on both the uplink and the downlink, which increase the correlation between the outage events experienced on the two links.

B. Cognitive Channel Assignment

In *cognitive channel assignment* (CCA), the base station employs spectrum sensing with respect to subcarriers in WiMAX. To achieve this, the BS must be equipped with ultra-sensitive signal detectors [31], [35], [36]. In the proposed CCA scheme, the cognitive radio functions of spectrum sensing and allocation is performed at the BS. Subchannels allocated to a call are simply deallocated to the general pool when the call ends. Thus, the MS does not need to be modified relative to conventional WiMAX. The use of power control in the CCA scheme minimizes cochannel interference and maximizes frequency reuse. The CCA technique can also be applied to FDMA-based systems. In this case, the unit of bandwidth allocation is a frequency channel rather than an OFDMA subcarrier.

Consider a WiMAX system with a set \mathcal{S} of subcarriers available for OFDM data transmission. As discussed earlier, the OFDMA scheme of WiMAX allows multiple connections to share a given subcarrier within a TDM frame. However, once a subcarrier is allocated to a base station, suballocation of slots within the TDM frame can be managed locally by the base station without the need to consider cochannel interference constraints. Thus, without loss of generality, we shall assume that the subcarrier is the smallest unit of bandwidth that can be allocated to a connection or call.

1) *Frequency-Division Duplexing*: If frequency-division duplexing (FDD) is used, the set \mathcal{S} can be partitioned into two sets \mathcal{U} and \mathcal{D} (see Fig. 1(b)):

$$\mathcal{S} = \mathcal{U} \cup \mathcal{D}, \quad (20)$$

where \mathcal{U} is the set of subcarriers reserved for uplink transmissions and \mathcal{D} is the set of subcarriers set aside for downlink transmissions. Let \mathcal{U}_i and \mathcal{D}_i denote, respectively, the set of uplink and downlink subcarriers currently in use by BS_i .

When a new call request from an MS arrives to cell i , the serving base station BS_i examines the received power levels

from each of the subcarriers and determines the set $\hat{\mathcal{U}}_i \subseteq \mathcal{U} \setminus \mathcal{U}_i$ of *free* uplink subcarriers as follows:

$$\hat{\mathcal{U}}_i = \bigcup \{c \in \mathcal{U} \setminus \mathcal{U}_i : R_c < \eta_{\text{up}}\}, \quad (21)$$

where R_c is the received signal power from subcarrier c and η_{up} is a detection threshold associated with uplink communications. Thus, $\hat{\mathcal{U}}_i$ is the set of subcarriers not already in use by BS_i with received power less than the detection threshold η_{up} . Similarly, the base station determines the set $\hat{\mathcal{D}}_i \subseteq \mathcal{D} \setminus \mathcal{D}_i$ of *free* downlink subcarriers:

$$\hat{\mathcal{D}}_i = \bigcup \{c \in \mathcal{D} \setminus \mathcal{D}_i : R_c < \eta_{\text{down}}\}, \quad (22)$$

where η_{down} is a detection threshold associated with downlink communications. The determination of η_{up} and η_{down} are discussed later in this section.

The base station selects a set of subcarriers from the set $\hat{\mathcal{U}}_i$ to form an uplink subchannel C_{up} to satisfy the bandwidth requirements of the call request on the uplink. If an insufficient number of uplink subcarriers is free, the call is blocked. Otherwise, the base station proceeds to allocate the downlink subchannel C_{down} by choosing a set of subcarriers from the set $\hat{\mathcal{D}}_i$ to meet the requirements of the call on the downlink. If an insufficient number of downlink subcarriers is free, the call is blocked. Then the sets \mathcal{U}_i and \mathcal{D}_i are updated as follows:

$$\mathcal{U}_i \leftarrow \mathcal{U}_i \cup C_{\text{up}}, \quad \mathcal{D}_i \leftarrow \mathcal{D}_i \cup C_{\text{down}}. \quad (23)$$

2) *Time-Division Duplexing*: If TDD is used, each subcarrier is used for both uplink and downlink transmissions in different portions of the TDM frame, as illustrated in Fig. 1(a). Similar to the FDD case, let \mathcal{U}_i and \mathcal{D}_i denote, respectively, the set of uplink and downlink subcarriers currently in use by BS_i .

When a new call request from an MS arrives to cell i , the serving base station BS_i examines the received power levels from each of the subcarriers. Each BS is equipped with a special *ultra-sensitive* signal detector [31], [35], [36], which estimates the received signal strength across all frequencies in the band of interest. In essence, the signal detector takes time series measurements and performs spectral estimation in the band of interest¹. The base station then determines the set $\hat{\mathcal{U}}_i \subseteq \mathcal{U} \setminus \mathcal{U}_i$ of *free* uplink subcarriers as follows (cf. (21)):

$$\hat{\mathcal{U}}_i = \bigcup \{c \in \mathcal{S} \setminus \mathcal{U}_i : R_c^{(u)} < \eta_{\text{up}}\}, \quad (24)$$

where $R_c^{(u)}$ is the received signal power from subcarrier c during the uplink portion of the TDM frame and η_{up} is the uplink detection threshold. Thus, $\hat{\mathcal{U}}_i$ is the set of uplink subcarriers not already in use by BS_i with received power on the uplink portion of the TDM frame less than the detection threshold η_{up} . Similarly, the base station determines the set $\hat{\mathcal{D}}_i \subseteq \mathcal{D} \setminus \mathcal{D}_i$ of *free* downlink subcarriers as follows (cf. (22)):

$$\hat{\mathcal{D}}_i = \bigcup \{c \in \mathcal{S} \setminus \mathcal{D}_i : R_c^{(d)} < \eta_{\text{down}}\}, \quad (25)$$

where $R_c^{(d)}$ is the received signal power from subcarrier c during the downlink portion of the TDM frame and η_{down} is the downlink detection threshold.

¹Such detectors have been built by the Shared Spectrum Company.

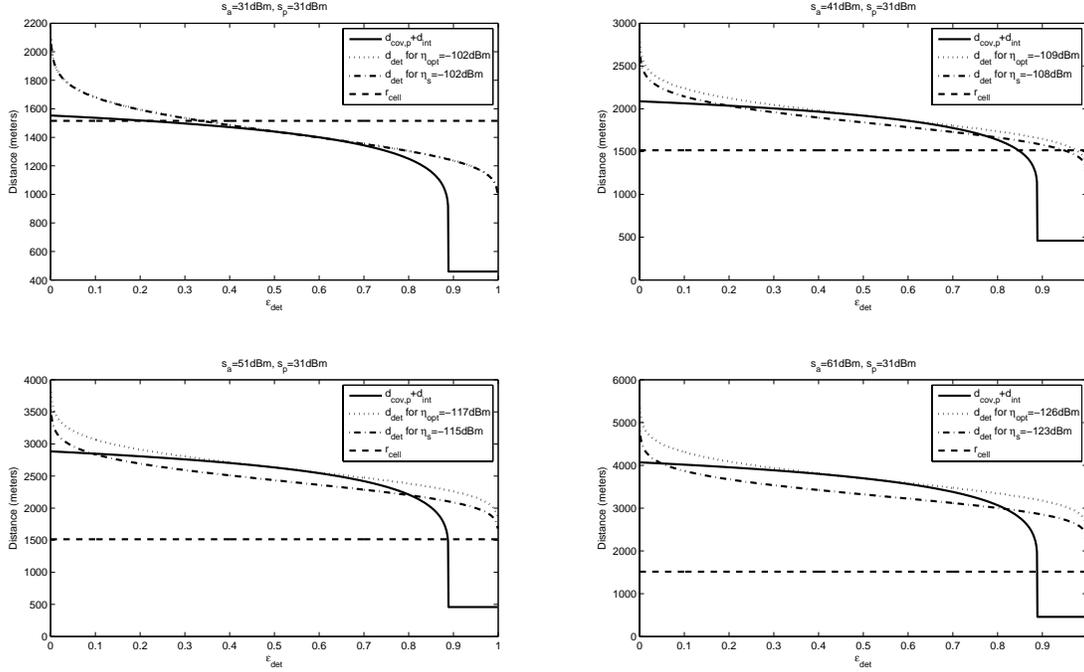


Fig. 7. Interference, coverage, and detection distances for downlink.

Using (17), we can then derive the condition

$$\epsilon_{\text{det}} < 1 - \left(\frac{\epsilon_{\text{int}}}{1 - P_{\text{out}}} \right) \cdot \left[1 - Q \left(\frac{s_a - i_{\text{max}}}{\sigma_{a,v}} \right) \right]^{-1}. \quad (31)$$

For the scenarios represented in Fig. 7, $\epsilon_{\text{int}} = 0.1$, $P_{\text{out}} = 0.1$, $i_{\text{max}} = -132$ dBm (cf. (34)), and $\sigma_{a,v} = 3$ dB. For any value of s_a in the range 31 to 61 dBm, the term in the square brackets in (31) is approximately one and hence the right-hand side of (31) is approximately 0.89. Hence, as $\epsilon_{\text{det}} \rightarrow 0.89$, $d_{\text{int}} \rightarrow 0$ for the four values of s_a considered in Fig. 7 and the solid curve tends to the value $d_{\text{cov,p}}$.

The system parameters for the downlink scenario are as follows:

- CR downlink transmitter power $s_a^{(d)}$ varies from 31 to 61 dBm;
- primary transmitter power s_p varies from 31 to 61 dBm;
- system bandwidth is 10 MHz and the number of sub-channels is 16;
- receiver noise figure $\text{NF} = 8$ dB, the required signal-to-noise ratio $\text{SNR} = 0.8$ dB, and the receiver gain is $\text{RG} = 0$ dB;
- interference probability $P_{\text{int}} = 0.01$;
- maximum outage probability $P_{\text{out}} = 0.1$;
- minimum received signal threshold at the subscriber station:

$$r_{\text{min}} = N_{\text{floor}} + \text{NF} + \text{SNR} - \text{RG}, \quad (32)$$

where the noise floor is given by

$$N_{\text{floor}} = -174 \text{ dBm/Hz} + 10 \log_{10} B = -134 \text{ dBm}, \quad (33)$$

where the subcarrier bandwidth is $B = 10$ kHz.

- maximum interference threshold at the subscriber station:

$$i_{\text{max}} = N_{\text{floor}} + \text{NF} + \text{INR}, \quad (34)$$

where N_{floor} and NF have values given above. In the numerical results presented in Section V, we have set $\text{INR} = -6$ dB.

The downlink antenna parameters are as follows:

- carrier frequency $f_c = 2.4$ GHz;
- victim receiver (MS) antenna height, $h_v = 2$ m ;
- CR and primary transmitter (BS) antenna height, $h_a = h_p = 10$ m.

The parameter values selected above reflect worst-case scenarios within the following range of parameter values, for which the Erceg models are valid [2]:

$$1.9 \text{ GHz} \leq f \leq 3.5 \text{ GHz},$$

$$10 \text{ m} \leq h_b \leq 80 \text{ m},$$

$$2 \text{ m} \leq h_m \leq 10 \text{ m},$$

$$0.1 \text{ km} \leq d \leq 8 \text{ km}.$$

The propagation loss between the transmitters and receivers is assumed to follow the Erceg C path loss model [33] with shadowing standard deviations $\sigma_{a,v} = \sigma_{p,v} = 6$ dB and $\sigma_{a,p} = 3$ dB. The Erceg model is commonly used in performance studies of WiMAX [2].

Observe that both the detection distance and the interference distance are monotonically decreasing functions of the detection probability. To avoid harmful interference, the detection threshold should be chosen such that the noninterference (27)

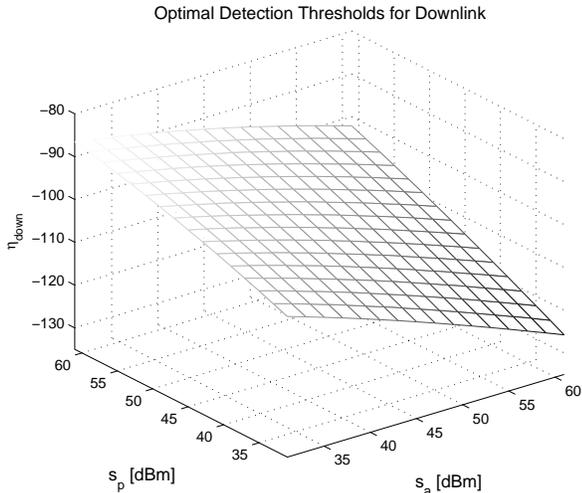


Fig. 8. Optimal detection thresholds for downlink transmissions.

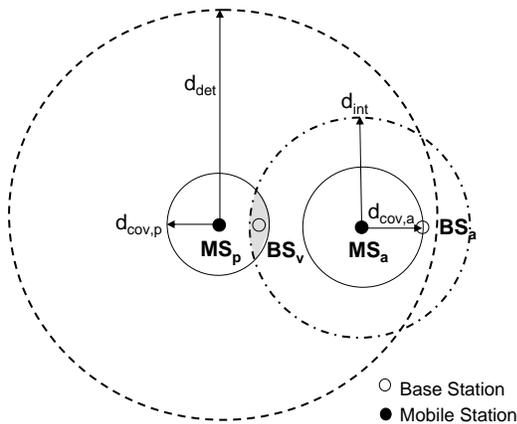


Fig. 9. Interference condition for uplink.

condition holds over the entire range of values of ϵ_{det} , since we have no prior knowledge of the detection probability. We consider the power control range for the downlink transmission between 31 and 61 dBm. Fig. 8 shows the optimal detection thresholds for the LBT algorithm as surfaces for any combination of the transmit power levels in the range specified above at other base stations.

The optimum LBT downlink detection threshold ranges from $\eta_{\text{opt}} = -125$ to -102 dBm, while the simplified formula gives a suboptimal LBT threshold value ranging from $\eta_s = -123$ to -102 dBm. As discussed in [36], SSC detectors have been demonstrated to achieve a sensitivity as low as 15 dB below thermal noise. For a 625 kHz bandwidth, this is a sensitivity of approximately -125 dBm. Thus, downlink power allocation for cognitive WiMAX can be realized by modern ultra-sensitive detector technology.

D. Uplink Allocation

Fig. 9 presents the LBT sensing scenario for uplink transmission, where the primary node, MS_p , is now a subscriber station and the frequency-agile node, BS_a , is a base station.

The victim node, BS_v , is a base station situated within the coverage region around the primary subscriber, MS_p . The uplink transmission power allocation, $s_a^{(u)}$, for the new subscriber station MS_a is based on the LBT algorithm and power control employed at its receiving base station BS_a .

Similarly as for downlink transmission, we determine the detection and interference distances for base station and subscriber station, respectively, to avoid co-channel interference on uplink transmissions. Similar to the noninterference condition (27) on the downlink, a simple sufficient condition for the victim subscriber not to suffer harmful interference can be expressed by an inequality relating the coverage distances $d_{\text{cov},p}$, $d_{\text{cov},a}$, the detection distance d_{det} , and the interference distance d_{int} as follows:

$$d_{\text{int}} \leq d_{\text{det}} - d_{\text{cov},p} - d_{\text{cov},a}. \quad (35)$$

For uplink transmissions, the receivers are the base stations that are located at the centers of their associated cells. A set of subcarriers can be assigned to a user for the uplink connection with the base station in its cell as long as this does not cause cochannel interference to neighboring base stations that are currently tuned to some of the subcarriers in the same set. Fig. 10 illustrates three scenarios of the uplink spectrum allocation with a subscriber station MS_0 situated in different locations within cell 0:

- **Case a:** When MS_0 is located close to BS_0 , the interference range does not include any neighboring base stations. In this case any of the subcarriers in the assigned set can be reused in any of the surrounding cells for uplink connections.
- **Case b:** When MS_0 is located close to the edge delimiting two neighboring cells, the interference range includes a neighboring base station. In this case any of the subcarriers in the assigned set can be reused by any of the neighboring base stations except BS_4 .
- **Case c:** When MS_0 is located at the borders of three cells, the interference range can include three or more base stations. In this case any of the subcarriers in the assigned set can be reused by the neighboring base stations that are not included in the interference region, i.e., BS_1 , BS_2 , BS_3 , and BS_6 .

Similar to the downlink detection threshold given in (28), the optimum uplink LBT detection threshold can be found as follows:

$$\eta_{\text{up}} = s_p - \alpha_{p,a} - g \left(g^{-1}(s_a - i_{\text{max}} + \alpha_{a,v}) + g^{-1}(s_a - r_{\text{min}} + \alpha_{a,v}) + g^{-1}(s_p - r_{\text{min}} + \alpha_{p,v}) \right). \quad (36)$$

In this case, we also denote s_a by $s_a^{(u)}$ to emphasize that it represents the CR node transmit power on the uplink. Furthermore, a (suboptimal) approximate uplink LBT detection threshold can be obtained from (36) by setting $\alpha_{p,a} = \alpha_{a,v} = \alpha_{p,v} = 0$.

We consider an uplink scenario with the following system parameters:

- CR transmitter power $s_a^{(u)}$ varies from -3 to 27 dBm;
- primary transmitter power s_p varies from -3 to 27 dBm;

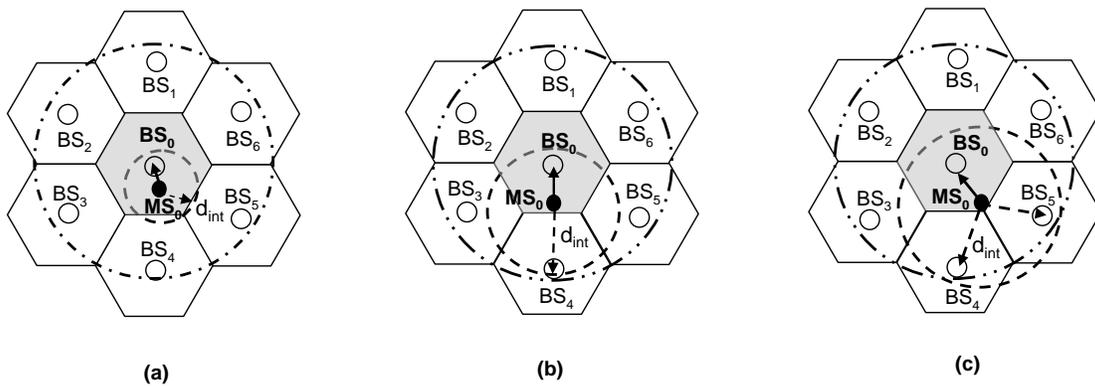


Fig. 10. Uplink spectrum occupancy map.

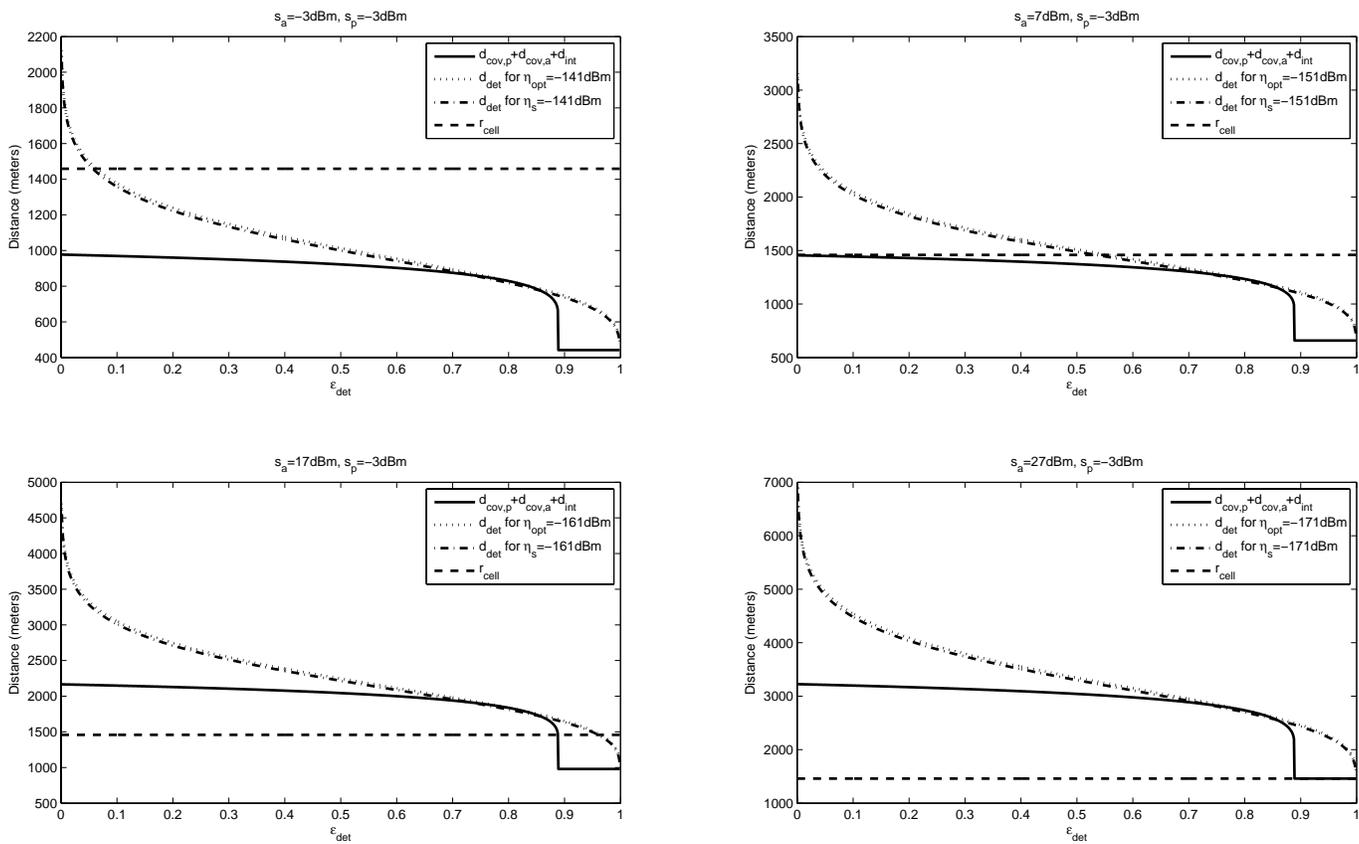


Fig. 11. Interference, coverage, and detection distances for uplink.

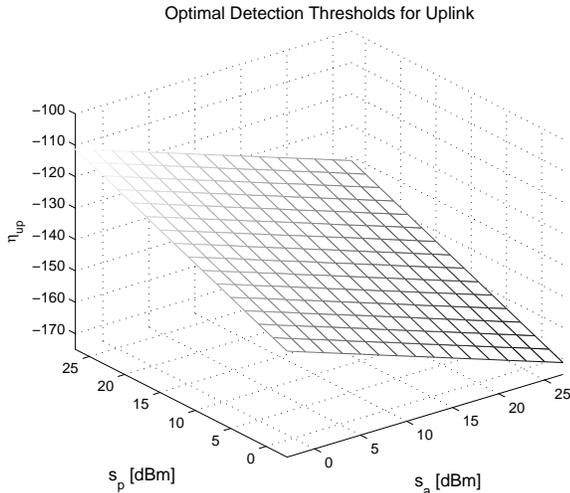


Fig. 12. Optimal detection thresholds for uplink transmissions.

- interference probability $P_{\text{int}} = 0.01$;
- maximum outage probability $P_{\text{out}} = 0.1$;
- minimum received signal threshold at the BS is given by (32) where the noise figure is $NF = 4$ dB, the signal-to-noise ratio is $SNR = 1.8$ dB, and the receiver gain is $RG = 18$ dB. The noise floor is as given in (33);
- maximum interference threshold at the BS is given by (34) where the noise figure NF and noise floor are as given above. For the interference-to-noise ratio INR , we have used values of 6 and -6 dB in our numerical results (see Section V).

The uplink antenna parameters are:

- CR and victim receiver antenna height, $h_v = 10$ m;
- CR and primary transmitter antenna height, $h_a, h_p = 2$ m.

The propagation loss between the transmitters and receivers is given by the Erceg C path loss model [33] with shadowing standard deviations $\sigma_{a,v} = \sigma_{p,v} = \sigma_{a,p} = 6$ dB.

With respect to Fig. 11, the noninterference condition requires the d_{det} curve to lie above the $d_{\text{int}} + d_{\text{cov},p} + d_{\text{cov},a}$ curve for the optimum LBT detection thresholds (cf. (36)). The power control range for uplink transmission is assumed to lie between -3 and 27 dBm. Fig. 12 shows the optimal uplink detection thresholds for the LBT algorithm as surfaces for any combination of transmit power levels at subscriber stations and agile nodes. The behavior of the curves representing $d_{\text{cov},p} + d_{\text{int}}$ can be derived analytically as discussed in Section IV-D.

The optimal uplink detection threshold ranges from $\eta_{\text{opt}} = -169$ to -147 dBm, while the simplified formula gives a suboptimal LBT threshold ranging from $\eta_s = -170$ to -149 dBm. Detectors with sensitivities as low as -169 dBm are not physically realizable with current technology [36]. Uplink power allocation requires greater detection sensitivity than downlink allocation primarily due to the lower transmit power used by the subscriber station on the uplink.

To circumvent this problem, downlink detection can be leveraged to perform uplink power allocation by imposing the requirement that a whenever a subcarrier is assigned on the

uplink for a user, a portion of the downlink TDM subframe (see Fig. 1(a)) must be reserved for the same user. In this way, detection of a downlink signal for a given subcarrier implies that the uplink subcarrier is also occupied. Thus, power allocation on the uplink is determined via signal detection on the downlink.

V. PERFORMANCE STUDY

In this section, we study the performance of cognitive WiMAX with respect to call blocking as a quality-of-service metric. Our objective is to show that LBT-based cognitive channel assignment can result in significant increases in network capacity compared to the conventional FCA and DCA channel assignment schemes.

For simplicity, we assume that all subchannels are of equal bandwidth. Furthermore, we consider only *unsolicited grant services* (UGS) in the context of voice circuit emulation. Under the assumption of OFDMA subchannels of equal bandwidth, the channel assignment behavior of an OFDMA system reduces to that of a conventional FDMA-based cellular system. However, one should keep in mind that the OFDMA subchannels are still narrowband to alleviate the problem of frequency-selective fading. Furthermore, we shall consider only TDD-based systems. In this context, we make the standard assumptions [51]–[54] that call requests arrive according to a Poisson arrival process and that call holding durations are exponentially distributed. The cellular network is assumed to have a hexagonal layout. The spatial distribution of requesting subscriber stations is assumed to be uniform over the coverage area.

A. Blocking Probability Evaluation

The blocking probability for the FCA scheme (without channel borrowing) can be expressed in closed form by the Erlang loss formula (cf. [11]):

$$B_{\text{FCA}} = \frac{\rho^J / J!}{\sum_{j=0}^J \rho^j / j!}, \quad (37)$$

where J is the number of channels allocated to a given cell and ρ is the traffic intensity in the cell. The traffic intensity ρ is specified in units of Erlangs and can be expressed as λ/μ , where λ is the mean arrival rate to a cell and $1/\mu$ is the mean cell holding time. The FCA blocking probability given by (37) provides a useful lower bound on the performance of channel assignment schemes. An upper bound on performance can be obtained by considering the blocking probability when the frequency reuse factor [11] is unity, i.e., all of the system channels are allocated to each cell. If N is the cluster size, then the total number of channels in the system is $K = JN$. Hence, the blocking probability under unit reuse factor (RF) is given by

$$B_{\text{RF1}} = \frac{\rho^K / K!}{\sum_{k=0}^K \rho^k / k!}. \quad (38)$$

We evaluate the blocking probabilities for DCA and CCA using computer simulation. The simulation program generates call requests according to a Poisson process. When an arrival

event occurs, the location of the subscriber station is drawn from a uniform distribution over the coverage area. Based on the subscriber location, the corresponding cell is determined. The call holding time in the cell is drawn from an exponential distribution.

To simulate the operation of DCA, a list of channels is maintained for each base station in the system. Each channel is marked as either *free* or *occupied*. When a call request arrives to a cell i , the call is assigned a free channel, if one is available. In this case, the channel is marked as occupied for all other cells in the cluster centered at cell i . Otherwise, if no free channel is available, the call is blocked. The simulation program performs the appropriate updates of the base station channel lists whenever a channel is released by a call.

Similar to the case of DCA, simulation of CCA requires that lists of channels be maintained for each base station. However, the criteria for deciding when a channel is free or occupied in CCA is based on LBT detection and interference distances, rather than the cell cluster structure. Note that the cell cluster concept ensures that co-channel cells are separated by a fixed distance called the *co-channel distance* (cf. [11]). By contrast, CCA exploits spectrum sensing and power control to achieve greater frequency reuse than DCA.

B. Numerical Results

Next we present some representative performance results comparing the three channel assignment schemes FCA, DCA, and CCA. We choose to display our results in terms of the percentage of accepted customer connections, which is more commonly used than blocking probability in the evaluation of practical cellular systems. The percentage of accepted connections is simply one minus the blocking probability.

The cellular service area consists of a 10×10 layout of hexagonal cells. The arrival of a call is assumed to be a Poisson process with a variable mean arrival rate expressed in calls/second and fixed mean call departure rate expressed in seconds. The ratio between the mean arrival rate and the mean call departure rate is the traffic intensity or traffic load. We assume a two-tier cluster of hexagonal cells such that the frequency reuse factor is 19, i.e., the cluster size is $N = 19$. The radio system parameters are as given in Sections IV-C and IV-D.

Fig. 13 shows the system capacity in terms of the simultaneous customer connections when 95 channels are available throughout a range of traffic intensity from 0 to 190 Erlangs/cell. In the figure, the curves marked with squares and triangles correspond to the FCA and DCA schemes, respectively. The FCA curve provides a lower bound on the performance of the DCA and CCA schemes, and is obtained using (37). The three curves that are marked with pentagrams in each figure show that the system capacity of CCA increases as the probability of interference increases. The DCA and CCA curves were obtained by computer simulation. Each simulation result was obtained by averaging over 10,000 trials. In each case, the 95% confidence intervals are less than 1% in width and are omitted from the figures for the sake of clarity. The curve marked with circles shows the performance of FCA

with unit reuse factor, obtained using (38). This curve provides an upper bound on the performance of the CCA scheme. The gap between the CCA curves and the upper bound curve shows that there is room for even greater capacity improvement.

In Fig. 13, we observe that CCA can support increased traffic intensity with a factor of relative to 2 to 3 DCA such that more than 90% of the customer calls are simultaneously connected, as the probability of interference is varied from 0.001 to 0.1. Whereas the simulation results in Fig. 13 correspond to an Interference-to-Noise Ratio (INR) of 6 dB, Fig. 14 shows the system capacity when $\text{INR} = -6$ dB. In Fig. 14, we observe that the CCA can support increased traffic intensity with a factor of 1.3 to 2 relative to DCA such that 90% of the customer calls are simultaneously connected, as the probability of interference is varied from 0.001 to 0.1. Figs. 15 and 16 show that doubling the number of channels to 190, for $\text{INR} = 6$ dB and $\text{INR} = -6$ dB, respectively, does not significantly change the relative performance of CCA vs. DCA.

VI. CONCLUSION

We proposed a framework for applying frequency agile cognitive radio technology to WiMAX networks. The proposed cognitive WiMAX architecture involves equipping the base stations with frequency agile cognitive radios with ultra-sensitive signal detectors. The base stations employ a cognitive channel assignment (CCA) scheme to allocate subcarriers. The CCA scheme combines power control, spectrum sensing, and Listen-Before-Talk dynamic spectrum access to achieve a substantial gain in frequency reuse relative to the conventional Fixed Channel Assignment (FCA) and Dynamic Channel Assignment (DCA) schemes.

We presented numerical results comparing the performance of CCA, DCA, and FCA in a simplified cellular network scenario of circuit-switched call requests arriving according to a Poisson process with exponential call holding times. Subscriber stations are spatially distributed according to a uniform distribution over a cellular service area with hexagonal cell layout. Our numerical results that showed that the CCA scheme could achieve from 30-300% higher capacity compared to DCA.

Whereas DCA requires the coordination of base station channel allocations using mobile switching centers (MSCs), CCA does not require coordination between base stations. However, CCA does require the use of power control and sensitive signal detectors at the base stations. Power control is a relatively mature technology that is employed in many practical cellular systems. Assuming that power control is available, the use of CCA can simplify network operations with respect to DCA, at the expense of employing cognitive radio technology to the base stations. We remark that if DCA is similarly enabled with power control and the CCA observations, the achievable capacity of DCA should be virtually indistinguishable from CCA capacity.

Our numerical results considered circuit-switched voice traffic, whereas WiMAX can support a rich set of traffic services with varying bandwidth requirements. To model variable bit rate traffic, more sophisticated packet-level traffic

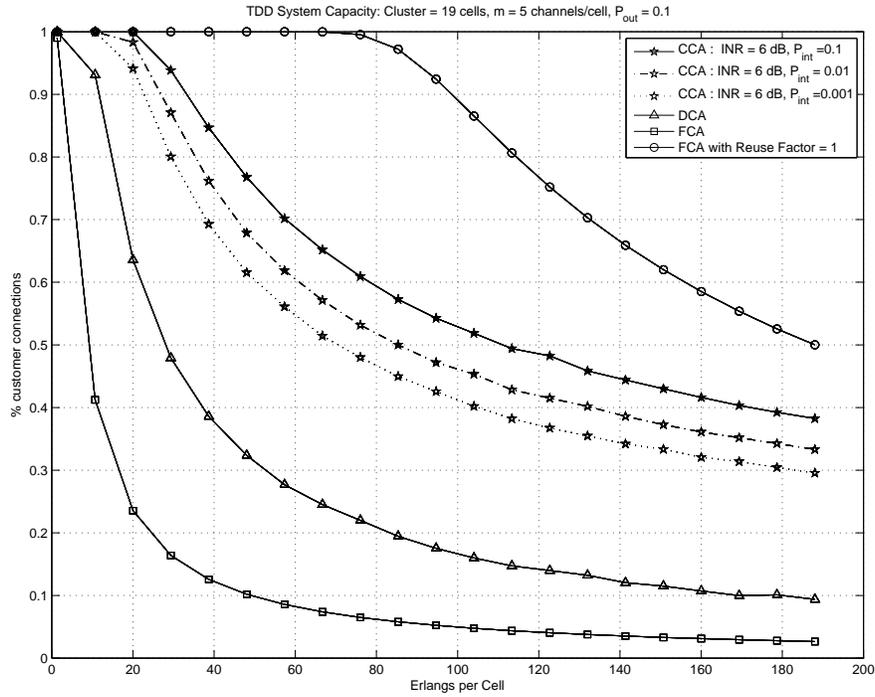


Fig. 13. TDD system capacity: CCA vs. DCA vs. FCA, $m = 95$ channels, INR = 6 dB.

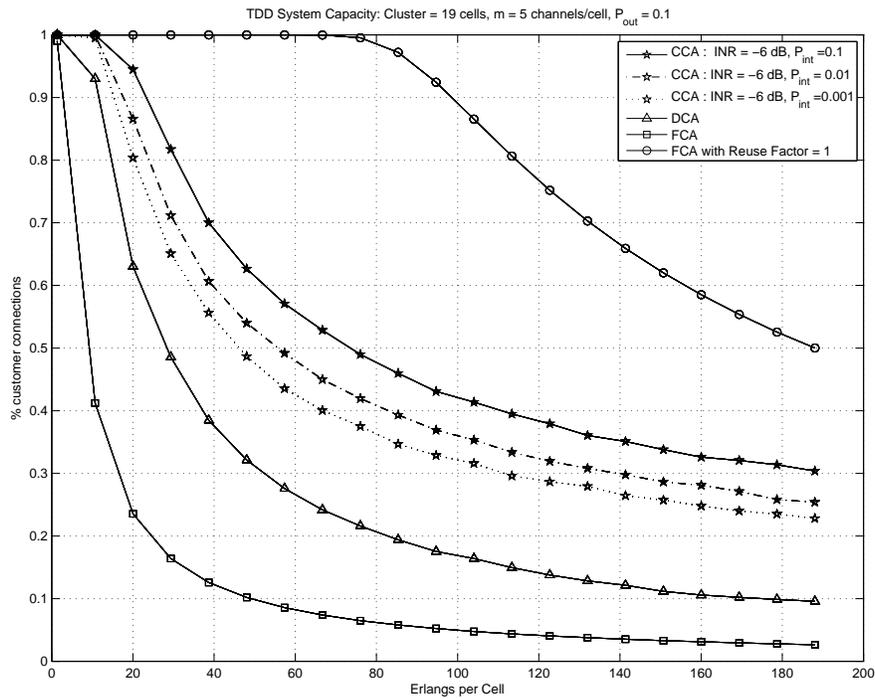


Fig. 14. TDD system capacity: CCA vs. DCA vs. FCA, $m = 95$ channels, INR = -6 dB.

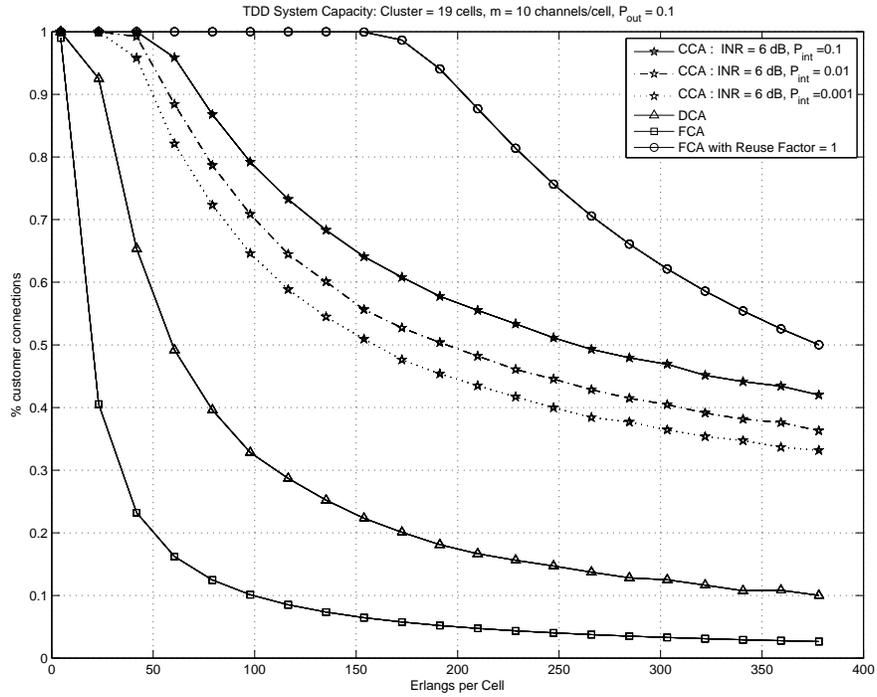


Fig. 15. TDD system capacity: CCA vs. DCA vs. FCA, $m = 190$ channels, INR = 6 dB.

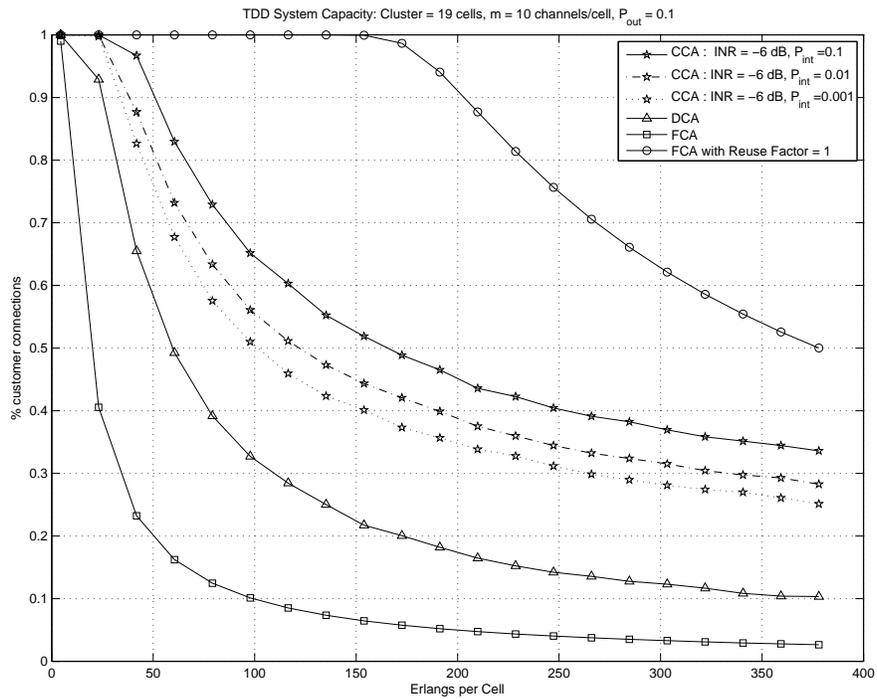


Fig. 16. TDD system capacity: CCA vs. DCA vs. FCA, $m = 190$ channels, INR = -6 dB.

arrival processes should be incorporated into the simulation of cognitive WiMAX. Also, the simulation program did not model the impact of user mobility. It would be of interest to evaluate cognitive WiMAX performance using realistic user mobility patterns.

The cognitive WiMAX architecture presented in this paper employs relatively simple frequency-agile radio techniques. Our objective was to show that substantial capacity improvements can be achieved via a practically realizable architecture using basic cognitive radio technology. The cognitive channel assignment (CCA) scheme proposed in this paper employs a simple LBT spectrum access scheme performed at the base station. As shown in the numerical results, there is still a significant gap between the performance of CCA and a system with unit reuse factor. In [3], it was shown that the use of collaboration can significantly improve dynamic spectrum access performance. More sophisticated CCA schemes could involve collaborative sensing among base stations and subscriber stations. Employing cognitive radio technology in the subscriber stations would enable spectrum sensing on the uplink channel with realizable detectors, provided a suitable subscriber-to-subscriber propagation model were available. The introduction of cognitive radio technology in the subscriber station opens up the possibility of multi-hop packet forwarding, which could further increase frequency reuse in the network, at the expense of higher cost on the subscriber side.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers, M. Aregbesola, and P. Tenhula for comments that have helped to improve the presentation of the paper.

REFERENCES

- [1] C. Eklund, R. Marks, S. Ponnuswamy, K. Stanwood, and N. van Wales, *Wireless MAN: Inside the IEEE 802.16 Standard for Wireless Metropolitan Area Networks*. IEEE Press, May 2006.
- [2] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX*. Prentice Hall, 2007.
- [3] A. E. Leu, M. McHenry, and B. L. Mark, "Modeling and Analysis of Interference in Listen-Before-Talk Spectrum Access Schemes," *Int. J. Network Mgmt.*, vol. 16, pp. 131–147, 2006.
- [4] A. Jovicic and P. Viswanath, "Cognitive radio – an information-theoretic perspective," in *Proc. Int. Symp. on Info. Theory (ISIT)*, (Seattle, WA), pp. 2413 – 2417, July 2006.
- [5] X. Zhu, L. Shen, and T.-S. P. Yum, "Analysis of cognitive radio spectrum access with optimal channel reservation," *IEEE Comm. Letters*, vol. 11, pp. 304 – 306, April 2007.
- [6] T. Yucek and H. Arslan, "Spectrum characterization for opportunistic cognitive radio systems," in *Proc. IEEE Milcom*, pp. 1–6, Oct. 2006.
- [7] M. E. Sahin and H. Arslan, "System design for cognitive radio communications," in *Proc. IEEE CROWNCOM*, June 2006.
- [8] WiMAX Forum, "Mobile WiMAX – Part I: A technical overview and performance evaluation." March 2006.
- [9] WiMAX Forum, "Mobile WiMAX – Part II: A comparative analysis," Mar. 2006.
- [10] J. P. Conti, "The Long Road to WiMAX (Wireless MAN Standard)," *IEE Review*, vol. 51, pp. 38–42, Oct. 2005.
- [11] J. W. Mark and W. Zhuang, *Wireless Communications and Networking*. Prentice Hall, 2003.
- [12] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Prentice Hall, 2002.
- [13] IEEE, "Standard 802.16-2004. Part 16: Air interface for fixed broadband wireless access systems," Oct. 2004.
- [14] IEEE, "Standard 802.16e-2005. Amendment to IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems- Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands," Dec. 2005.
- [15] IEEE, "Standard 802.16c-2002. Part 16: Air Interface for Fixed Broadband Wireless Access Systems – Amendment 1: Detailed System Profiles for 1066 GHz," Jan. 2003.
- [16] M. Morelli, C.-C. J. Kuo, and M.-O. Pun, "Synchronization Techniques for Orthogonal Division Multiple Access (OFDMA): A Tutorial Review," *Proceedings of IEEE*, vol. 95, July 2007.
- [17] H. Wang, W. Li, and D. P. Agrawal, "Dynamic Admission Control and QoS for 802.11 Wireless MAN," in *Proc. Wireless Telecommunication Symposium '05*, pp. 60–66, Apr. 2005.
- [18] T. J. Kahwa and N. Georganas, "A hybrid channel assignment scheme in large scale cellular-structured mobile communication systems," *IEEE Trans. on Commun.*, vol. COMM 26, pp. 432 – 438, 1978.
- [19] J. Sin and N. Georganas, "A simulation study of a hybrid channel assignment scheme for cellular land-mobile radio systems with erlang-c service," *IEEE Trans. on Commun.*, vol. COMM-9, pp. 143 – 147, 1981.
- [20] W. C. Y. Lee, *Mobile Cellular Telecommunications: Analog and Digital Systems*. McGraw Hill, 1995.
- [21] D. Cox, , and D. O. Reudink, "Increasing channel occupancy in large scale mobile radio systems: Dynamic channel reassignment," *IEEE Trans. on Veh. Technol.*, vol. 22, no. 4, pp. 218 – 222, 1973.
- [22] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems - a comprehensive survey," *IEEE Pers. Comm.*, pp. 10–31, June 1996.
- [23] X. Fung, C. Zhu, and P. Fun, "Greedy-based dynamical channel assignment strategy for cellular mobile networks," *IEEE Comm. Letters*, vol. 4, July 2000.
- [24] W. C. Y. Lee, "New cellular schemes for spectral efficiency," *IEEE Trans. on Veh. Technol.*, vol. VT-6, pp. 188 – 192, 1987.
- [25] A. J. Rustako et al., "Radio propagation measurements at microwave frequencies for microcellular mobile personal communications," *IEEE Trans. on Veh. Technol.*, vol. 2, pp. 203 – 210, 1991.
- [26] M. Frodigh, "Reuse partitioning combined with traffic adaptive channel assignment for highway microcellular radio systems," in *Proc. IEEE Globecom*, (Orlando, FL), pp. 1414–1418, Dec. 1992.
- [27] C.-L. I and P.-H. Chao, "Local packing - distributed dynamic channel allocation at cellular base station," in *Proc. IEEE Globecom*, (Houston, TX), pp. 293 – 301, Nov. 1993.
- [28] C.-L. I and P.-H. Chao, "Distributed dynamic channel allocation algorithms with adjacent channel constraints," in *Proc. IEEE Globecom*, vol. 1, pp. 169 – 175, Sept. 1994.
- [29] M. Serizawa and D. J. Goodman, "Instability and deadlock of distributed dynamic channel allocation," in *Proc. 43rd IEEE Vehicul. Tech. Conf.*, vol. 1, pp. 528 – 531, May 1993.
- [30] Y. Akaiwa and H. Andoh, "Channel segregation-a self organized dynamic allocation method: Application to TDMA/FDMA microcellular system," *IEEE J. Selected Areas in Comm.*, vol. 11, pp. 949 – 954, Aug. 1993.
- [31] A. E. Leu, K. Steadman, M. McHenry, and J. Bates, "Ultra Sensitive TV Detector Measurements," in *Proc. IEEE Int. Symp. on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, pp. 30–36, Nov. 2005.
- [32] M. McHenry, "The Probe Spectrum Access Method," in *Proc. IEEE Int. Symp. on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, pp. 346–351, Nov. 2005.
- [33] V. Erceg et al., "An empirically based path loss model for wireless channels in suburban environments," *IEEE J. on Selected Areas in Commun.*, vol. 17, pp. 1205–1211, June 1999.
- [34] M. N. Lustgarten and J. A. Madsen, "An Empirical Propagation Model (EPM-73)," *IEEE Trans. on Electromagnetic Compatibility*, vol. 19, Aug. 1977.
- [35] T. Erpek, A. E. Leu, and B. L. Mark, "Spectrum Sensing Performance in TV Bands using the Multitaper Method," in *Proc. IEEE 15th Signal Processing and Communication Applications Conference*, (Eskicehir, Turkey), June 2007.
- [36] K. D. Steadman, A. D. Rose, and T. N. Nguyen, "Dynamic spectrum sharing detectors," in *Proc. IEEE DySPAN*, (Dublin, Ireland), Apr. 2007.
- [37] C.-T. Chou, S. Shankar N, H. Kim, and K. G. Shin, "What and how much to gain by spectrum agility," *IEEE J. Selected Areas in Comm.*, vol. 25, pp. 576 – 588, Apr. 2007.
- [38] 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. Selected Areas in Comm.*, vol. 25, pp. 589 – 600, Apr. 2007.

- [39] J. C.-I. Chuang, "Performance issues and algorithms for dynamic channel assignment," *IEEE J. Selected Areas in Comm.*, vol. 11, pp. 955 – 963, Aug. 1993.
- [40] J. Zander, "Asymptotic bounds on the performance of a class of dynamic channel assignment algorithms," *IEEE Trans. on Vehicular Technology*, vol. 41, pp. 305 – 311, Aug. 1992.
- [41] J. Zander and H. Eriksson, "Asymptotic bounds on the performance of a class of dynamic channel assignment algorithms," *IEEE J. Selected Areas in Comm.*, vol. 11, pp. 926 – 933, Aug. 1993.
- [42] A. Dang, S. Zhu, and J. Cheng, "Globally optimized dynamic channel assignment for cellular radio systems," in *Proc. Int. Conf. on Comm. Technology (ICCT)*, vol. 2, Oct. 1998.
- [43] K. Smith and M. Palaniswami, "Static and dynamic channel assignment using neural networks," *IEEE J. Selected Areas in Comm.*, vol. 15, pp. 238 – 249, Feb. 1997.
- [44] H. Wang, W. Li, and D. P. Agrawal, "Dynamic channel allocation method using ANN and heuristic adjustment," in *Proc. of Neural Networks: IEEE World Congress on Computational Intelligence '98*, vol. 2, pp. 1624–1628, May 1998.
- [45] N. Weicker, G. Szabo, K. Weicker, and P. Widmayer, "Evolutionary multiobjective optimization for base station transmitter placement with frequency assignment," *IEEE Trans. on Evolutionary Computation*, vol. 7, pp. 189 – 203, Apr. 2003.
- [46] F. Borgonovo, A. Capone, and A. Molinaro, "The impact of signal strength measures on the efficiency of dynamic channel allocation techniques," in *Proc. IEEE Int. Conf. on Comm. (ICC)*, vol. 3, pp. 1400 – 1404, 1998.
- [47] N. Bambos, S. Chen, and G. Pottie, "Radio link admission algorithms for wireless networks with power control and active link quality protection," in *Proc. Infocom*, vol. 1, pp. 97 – 104, Apr. 1995.
- [48] K. Shoarinejad, J. L. Speyer, and G. J. Pottie, "Integrated predictive power control and dynamic channel assignment in mobile radio systems," *IEEE Trans. on Wireless Comm.*, vol. 2, pp. 976 – 988, Sept. 2003.
- [49] K. Madani and A. H. Aghvami, "Performance of distributed control channel allocation (DCCA) under uniform and non-uniform traffic conditions in microcellular radio communications," in *Proc. IEEE Int. Conf. on Comm. (ICC)*, vol. 1, pp. 206 – 210, May 1994.
- [50] K. A. West and G. L. Stüber, "An aggressive dynamic channel assignment strategy for a microcellular environment," *IEEE Trans. on Veh. Technol.*, vol. 43, no. 4, pp. 1027 – 1038, 1994.
- [51] ITU, "Guidelines for Evaluation of Radio Transmission Technologies for IMT-2000." Recommendation ITU-R M.1225, 1997.
- [52] D. Everitt and D. Mansfield, "Performance analysis of cellular mobile communication systems with dynamic channel assignment," *IEEE J. on Selected Areas in Commun.*, vol. 7, pp. 1172 – 1180, Oct. 1989.
- [53] M. Zhang and T. S. P. Yum, "Comparisons of channel assignment strategies in cellular mobile telephone systems," *IEEE Trans. on Vehicular Technology*, vol. 38, pp. 211 – 215, Nov. 1989.
- [54] M. Clark, V. Erceg, and L. J. Greenstein, "Reuse efficiency in urban microcellular networks," *IEEE Trans. on Veh. Technol.*, vol. 46, pp. 279–288, May 1997.

APPENDIX

A. Coverage, Detection, and Interference Distances

The signal strength received at node v from node p is given by

$$R_v = s_p - g(d_{p,v}) + W_{p,v}, \quad (39)$$

where $W_{p,v} \sim \mathcal{N}(0, \sigma_{p,v}^2)$. From (5), we have

$$\begin{aligned} P_{\text{out}} &= P\{s_p - g(d_{p,v}) + W_{p,v} < r_{\min} | E_p\} \\ &= P\{W_{p,v} < r_{\min} - s_p + g(d_{p,v}) | E_p\} \\ &= 1 - Q\left(\frac{r_{\min} - s_p + g(d_{p,v})}{\sigma_{p,v}}\right). \end{aligned} \quad (40)$$

By rearranging terms in (40), we obtain

$$d_{p,v} = g^{-1}(s_p - r_{\min} + \sigma_{p,v}Q^{-1}(1 - P_{\text{out}})). \quad (41)$$

Assuming that P_{out} is a continuous function of $d_{p,v}$, the definition of $d_{\text{cov},p}$ in (6) implies that $d_{p,v} = d_{\text{cov},p}$ when $P_{\text{out}} = \epsilon_{\text{out}}$, from which (7) follows:

$$d_{\text{cov},p} = g^{-1}(s_p - r_{\min} + \sigma_{p,v}Q^{-1}(1 - \epsilon_{\text{out}})).$$

The signal strength received at node a from node p is given by

$$R_a = s_a - g(d_{p,a}) + W_{p,a}, \quad (42)$$

where $W_{p,a} \sim \mathcal{N}(0, \sigma_{p,a}^2)$. From (9), we have

$$\begin{aligned} P_{\text{det}} &= P\{s_a - g(d_{p,a}) + W_{p,a} \geq \eta_{\text{det}}\} \\ &= P\{W_{p,a} \geq \eta_{\text{det}} - s_p + g(d_{p,a})\} \\ &= Q\left(\frac{\eta_{\text{det}} - s_p + g(d_{p,a})}{\sigma_{p,a}}\right). \end{aligned} \quad (43)$$

The definition of d_{det} in (10) implies that $d_{p,a} = d_{\text{det}}$ when $P_{\text{det}} = \epsilon_{\text{det}}$, from which (11) follows:

$$d_{\text{det}} = g^{-1}(s_p - \eta_{\text{det}} - \sigma_{p,a}Q^{-1}(1 - \epsilon_{\text{det}})).$$

From (14), we have

$$P_{\text{int}} = P\{I_v \geq i_{\max} | E_a\}P(E_a) \cdot P\{R_v \geq r_{\min} | E_p\}P(E_p). \quad (44)$$

We have

$$P\{I_v \geq i_{\max} | E_a\} = Q\left(\frac{i_{\max} + g(d_{a,v}) - s_a}{\sigma_{a,v}}\right) \quad (45)$$

$$P\{R_v \geq r_{\min} | E_p\} = 1 - P_{\text{out}}. \quad (46)$$

Now, assuming that the primary node p is always in the on state, the probability that node a is in the on state can be approximated as $P(E_a) \approx 1 - P_{\text{det}}$, assuming that the node a always has data to send. Hence, (44) can be written as

$$P_{\text{int}} = (1 - P_{\text{out}})(1 - P_{\text{det}})Q\left(\frac{i_{\max} + g(d_{a,v}) - s_a}{\sigma_{a,v}}\right). \quad (47)$$

Assuming that P_{int} is a continuous function of $d_{a,v}$, the definition of $d_{\text{cov},p}$ in (15) implies that $d_{a,v} = d_{\text{int}}$ when $P_{\text{int}} = \epsilon_{\text{int}}$, we obtain (16) follows:

$$d_{\text{int}} = g^{-1}(s_a - i_{\max} - \alpha_{a,v}),$$

where

$$\alpha_{a,v} = -\sigma_{a,v}Q^{-1}\left(\frac{\epsilon_{\text{int}}}{(1 - P_{\text{out}})(1 - P_{\text{det}})}\right).$$

Using the identity $Q^{-1}(x) \equiv -Q^{-1}(1 - x)$, we obtain (17).

B. Optimum LBT Detection

Replacing the terms in (19) with their mathematical formulas, as given in Section III-A we obtain the following inequality:

$$\begin{aligned} g^{-1}(s_p - \eta + \alpha_{p,a}) &\geq g^{-1}(s_a - i_{\max} - \alpha_{a,v}) \\ &\quad + g^{-1}(s_p - r_{\min} + \alpha_{p,v}). \end{aligned} \quad (48)$$

From (48), we determine that the LBT detection threshold must satisfy the following inequality:

$$\begin{aligned} \eta \leq & s_p - \alpha_{p,a} - g(g^{-1}(s_a - i_{\max} - \alpha_{a,v}) \\ & + g^{-1}(s_p - r_{\min} + \alpha_{p,v})). \end{aligned} \quad (49)$$

In particular, if we ignore the shadowing components, the LBT detection threshold satisfies

$$\eta \leq s_p - g(g^{-1}(s_a - i_{\max}) + g^{-1}(s_p - r_{\min})). \quad (50)$$