A Spectrum Decision Framework for Cognitive Radio Networks

Won-Yeol Lee, *Student Member*, *IEEE*, and Ian F. Akyildiz, *Fellow*, *IEEE*

Abstract—Cognitive radio networks have been proposed as a solution to both spectrum inefficiency and spectrum scarcity problems. However, they face to a unique challenge based on the fluctuating nature of heterogeneous spectrum bands as well as the diverse service requirements of various applications. In this paper, a spectrum decision framework is proposed to determine a set of spectrum bands by considering the application requirements as well as the dynamic nature of spectrum bands. To this end, first, each spectrum is characterized by jointly considering primary user activity and spectrum sensing operations. Based on this, a minimum variance-based spectrum decision is proposed for real-time applications, which minimizes the capacity variance of the decided spectrum bands subject to the capacity constraints. For best-effort applications, a maximum capacity-based spectrum decision is proposed where spectrum bands are decided to maximize the total network capacity. Moreover, a dynamic resource management scheme is developed to coordinate the spectrum decision adaptively dependent on the time-varying cognitive radio network capacity. Simulation results show that the proposed methods provide efficient bandwidth utilization while satisfying service requirements.

Index Terms—Cognitive radio networks, spectrum decision, spectrum characterization, real-time application, best-effort application, minimum variance-based spectrum decision, maximum capacity-based spectrum decision, resource management.

1 Introduction

Today's wireless networks are characterized as a static spectrum assignment policy. Recently, because of the increase in spectrum demand, this policy is faced with spectrum scarcity at particular spectrum bands. On the contrary, a large portion of the assigned spectrum is still used sporadically leading to underutilization of the significant amount of spectrum [9]. Hence, dynamic spectrum access techniques have recently been proposed to solve these spectrum inefficiency problems.

The key enabling technology for dynamic spectrum access techniques is the cognitive radio technology, which provides the capability to share the wireless channel with licensed users (or primary users) in an opportunistic manner [1]. Cognitive radio (CR) networks are envisioned to provide high bandwidth to mobile users via heterogeneous wireless architectures and dynamic spectrum access techniques. CR networks, however, impose unique challenges because of the high fluctuation in the available spectrum as well as the diverse quality-of-service (QoS) requirements of various applications. To address these challenges, first, CR networks are required to determine which portions of the spectrum are available, called *spectrum sensing* [2], [10]. Furthermore, how to coordinate multiple CR users to share the spectrum band, called *spectrum sharing*, is another important issue in CR networks [7], [16].

- *The authors are with the Broadband Wireless Networking Laboratory, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0250.*
E-mail: wylee@gatech.edu, ian@ece.gatech.edu.

Manuscript received 23 July 2008; revised 28 July 2009; accepted 24 Dec. 2009; published online 3 Aug. 2010.
For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number TMC-2008-07-0285.
Digital Object Identifier no. 10.1109/TMC.2010.147.

Although all these efforts enable CR users to exploit spectrum opportunities effectively, the heterogenous spectrum environment introduces a new critical issue in CR networks. Generally, CR networks have multiple available spectrum bands over a wide frequency range that show different channel characteristics, and need to support applications with diverse service requirements. Therefore, once available spectrum bands are identified through spectrum sensing, CR networks need to select the proper spectrum bands according to the application requirements. This process is referred to as *spectrum decision*, which constitutes an important but yet unexplored topic in CR networks. To decide on spectrum bands properly, CR networks need to consider the following issues:

- All available spectrum bands show different characteristics in the CR network. To select the proper spectrum, the CR network needs to characterize available spectrum bands by considering current radio conditions as well as the primary user (PU) activity.
- The CR network needs to provide a dynamic decision framework to consider all possible events that prevent reliable communications by closely interacting with other CR functionalities such as spectrum sensing and spectrum sharing.
- According to the PU activities, total capacity in CR networks varies over time, which makes it more difficult to decide on spectrum bands while maintaining the service quality of other CR users. Thus, the CR network should perform spectrum decision adaptively dependent on time-varying spectrum resources.

In this paper, an adaptive spectrum decision framework is proposed with the consideration of all decision events and application types. First, a novel capacity model is developed

Published by the IEEE CS, CASS, ComSoc, IES, & SPS

to describe unique characteristics in CR networks by considering PU activity as well as sensing capability. Accordingly, two different decision schemes are introduced. To satisfy the delay constraints in real-time applications, we propose a minimum variance-based spectrum decision (MVSD) scheme that selects spectrum bands to minimize capacity variation. For best-effort applications, we propose a maximum capacity-based spectrum decision (MCSD) scheme to maximize the total network capacity. Both decision schemes are controlled by a proposed resource management based on the current network condition.

The remainder of the paper is organized as follows: Section 2 presents previous research and our motivation. In Section 3, we propose a novel framework for spectrum decision. In Section 4, we present a spectrum capacity model used in this paper. Spectrum decision methods for real-time and best-effort applications are proposed in Sections 5 and 6, respectively. Then, we develop a dynamic resource management scheme in Section 7. Simulation results are presented in Section 8. Finally, conclusions are presented in Section 9.

2 Motivation

2.1 Related Work

The proposed spectrum decision has a similar objective to the spectrum sharing in the sense that it performs resource allocation based on service requirements. Most of the research on spectrum sharing in CR networks has mainly focused on how to efficiently allocate either spectrum or power among CR users subject to interference constraints.

For spectrum allocation, a global optimization scheme is developed based on graph theory [17]. However, whenever the network topology changes according to the node mobility, the network needs to completely recompute spectrum assignment leading to a higher computational and communication overhead. To solve this problem, a distributed spectrum allocation based on local bargaining is proposed in [4], where CR users negotiate spectrum assignment within local self-organized groups. For the resource-constrained networks such as sensor and ad hoc networks, a rule-based spectrum management is proposed, where CR users access the spectrum independently according to both local observation and predetermined rules [5]. In [20], a dynamic channel selection scheme is developed for delay-sensitive applications based on a priority queuing analysis and a decentralized learning algorithm.

Power allocation among CR users competing the same spectrum is another important issue in spectrum sharing. In [12], an optimal power allocation scheme is proposed to achieve ergodic and outage capacity of the fading channel under different types of power constraints and fading models. In [22], joint beam-forming and power allocation techniques are presented to maximize the user capacity while ensuring the QoS of primary users. Game theory provides an efficient distributed spectrum sharing scheme by describing the conflict and cooperation among CR users, and hence allowing each user to rationally decide on its best action. Thus, it has been widely exploited for both channel allocation [16] and for power allocation [7].

2.2 Implementation Challenge in Spectrum Decision

All of the previous research explained above has mainly addressed spectrum sharing issues where all operations are performed within the same spectrum band or across contiguous channels. Furthermore, to adapt the fast time-varying channels, they are generally designed as a short-term operation, such as a packet-based or a time-slot-based scheduling.

However, CR networks necessitate an additional resource allocation capability when primary users are detected or CR users newly begin their sessions, which are relatively long-term events. Thus, this capability should consider longer-term channel characteristics, compared to spectrum sharing. In addition, since available spectrum bands are distributed over a wide frequency range, this function needs to be implemented as an interspectrum operation. However, this operation inevitably introduces an additional switching delay leading to service quality degradation. Thus, it is not desirable to extend existing spectrum sharing solutions designed to adapt to the fast time-varying channel to the long-term interspectrum operation. This unique challenge in CR networks has not been addressed in previous research. Here our design objective of the spectrum decision framework is to decouple all inter-spectrum functionalities totally from spectrum sharing.

3 The Proposed Spectrum Decision Framework

3.1 System Model

In this paper, we consider an infrastructure-based CR network that has a centralized network entity, such as a base-station. The base-station exerts control over all CR users within its transmission range. CR users perform the observations and analysis on radio environments and feed them to the central base-station, which decides on spectrum availability and spectrum allocation. Each CR user has multiple software-defined radio (SDR) transceivers to exploit multiple spectrum bands over a wide frequency range by reconfiguring the operating frequency through software operations. Here, we assume frequency division duplex (FDD) systems where uplink and downlink channels are separated. Thus, the proposed decision scheme can be applied to each link independently.

When primary users appear in the spectrum band, CR users need to move to a new available band, resulting in a temporary communication break. To solve this problem, we assume that multiple noncontiguous spectrum bands can be simultaneously used for the transmission in the CR network. This method can create a signal that is not only capable of high data throughput, but is also immune to the PU activity. Even if a primary user appears in one of the current spectrum bands, the rest of them will maintain current transmissions [1].

The control channel plays an important role in exchanging information regarding sensing and resource allocation. Several methods are presented in [3], one of which is assumed to be used as the common control channel in our proposed method.

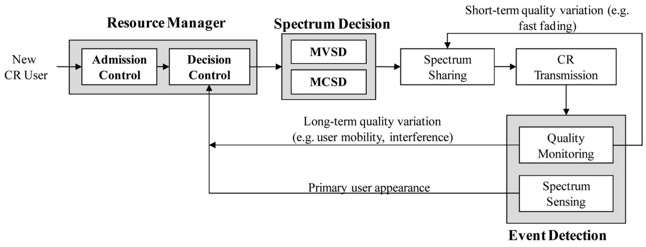


Fig. 1. The proposed spectrum decision framework.

3.2 Framework Overview

Based on the system model above, we develop a novel framework for spectrum decision. Here, spectrum decision is considered as an event-based functionality, i.e., the CR network decides on the proper spectrum bands in the following events:

- *CR user appearance:* When a new CR user appears in the CR network, it needs to be assigned to new spectrum bands for its transmission.
- *Primary user appearance:* When a primary user appears in the spectrum band, CR users should move to the new spectrum bands.
- *Channel quality degradation:* When channel condition becomes worse, CR users want to switch to a better spectrum band.

To consider all decision events effectively, the CR network necessitates a unified framework for spectrum decision. Fig. 1 shows the proposed framework for spectrum decision. A detailed description of this framework is as follows.

By considering current spectrum conditions, a *resource manager* determines if the CR network accepts a new incoming CR user or not. If a new CR user is allowed to transmit, it is assigned to the proper spectrum bands through *spectrum decision*. Since there may be multiple CR users competing the same spectrum, *spectrum sharing* coordinates those multiple accesses to prevent collisions, and accordingly to achieve the maximum capacity. In the *event detection*, current spectrum bands and users connections are monitored to detect decision events. The event detection consists of two main tasks: *spectrum sensing* and *quality monitoring*. When events are detected, the CR network reconfigures its resource allocation to maintain the service quality. In case of short-term channel variations such as fast fading, the CR network reallocates resources within the spectrum band through spectrum sharing. If a primary user is detected or the current spectrum band cannot provide the predetermined service quality any longer over a long-term period, the CR network switches the spectrum through the resource manager and the spectrum decision. In the proposed framework, CR users perform only event detection. Based on information gathered from CR users, the base-station decides on spectrum availability and performs spectrum decision as explained above.

Consequently, the proposed spectrum decision framework provides a hierarchical QoS guaranteeing scheme: spectrum sharing to allocate the channel and transmission power for short-term service qualities, and spectrum decision to determine the best spectrum for maintaining the service quality over a long term period. In this paper, we mainly focus on decision functionalities: *spectrum decision* and *resource management*. Spectrum sharing and event detection functionalities are out of the scope in this paper.

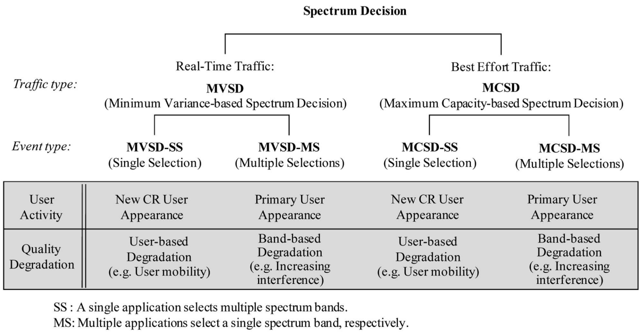


Fig. 2. Classification of the proposed spectrum decision.

3.3 Spectrum Decision Functionalities

In the proposed framework, we consider two types of applications: *real-time* and *best-effort* (in this paper, the terms "application" and "user" are interchangeably used). According to the application type, the proposed spectrum decision can be classified into a *minimum variance-based spectrum decision* for real-time applications, and a *maximum capacity-based spectrum decision* for best-effort applications.

Decision events mainly occur because of either user activities or quality degradations. When primary user appears in the spectrum band or a new CR user begins to transmit, the spectrum decision needs to be initiated. Moreover, the quality degradation of either an entire spectrum band (e.g., increase in interference) or a specific user connection (e.g., moving far from the base-station) can also trigger spectrum decision.

If a CR user exploits multiple spectrum bands, the spectrum decision method becomes more complicated according to the event type. When a new CR user appears or the QoS of a certain user becomes worse, multiple spectrum bands need to be determined for a single user at a time, called *single selection (SS)*. On the other hand, when a primary user appears or the quality of a certain spectrum band becomes worse, multiple CR users residing in that spectrum band lose one of their current spectrum bands, which requires multiple spectrum decisions for each CR user, called *multiple selections (MS)*.

As shown in Fig. 2, according to the traffic and event types, spectrum decision can be classified into four categories: MVSD-SS, MVSD-MS, MCSD-SS, and MCSD-MS, which are proposed in Sections 5.1, 5.2, 6.1, and 6.2, respectively.

For ease of representation, the important notations used in the subsequent discussion are summarized in Table 1.

4 Spectrum Characterization

To determine the spectrum band properly, it is important to identify the characteristics of each spectrum. To this end, in this section, we define the PU activity, and accordingly propose a novel CR capacity model.

TABLE 1
Symbols Used for the Analytical Modeling

Notations	Descriptions
N	Number of transceivers in CR users
$c_i(k)$	Normalized capacity of spectrum i at user k
$C_i^{\mathrm{CR}}(k)$	Normalized CR capacity of spectrum i at user k
α_i, β_i	Primary user departure and arrival rates in spectrum i
T_i^{off}	Expected transmission duration without switching at spectrum i
τ	Spectrum switching delay
η_i	Sensing efficiency
$R_{\mathrm{s}}(k)$	Sustainable rate at user k
$P_{\mathrm{loss}}^{\mathrm{th}}$	Target data loss rate
B_i	Total bandwidth of spectrum i
W_i	Currently available (idle) bandwidth of spectrum i
W_{av}	Total bandwidth currently available in the network
W_{R}	Total bandwidth currently used by real-time users,
W_{req}	Expected bandwidth required for spectrum decision
$W_{\min}$	Minimum bandwidth to satisfy the guaranteed QoS
ϵ, π, ρ	Operational thresholds for overload, outage probability, and balance

4.1 Primary User Activity

For an efficient spectrum utilization, the CR network needs to be aware of the traffic statistics of primary networks in each spectrum, called *PU activity*. The PU activity can be modeled as exponentially distributed interarrivals [21]. In this model, the PU activity in spectrum i is defined as a two-state birth-death process with death rate α_i and birth rate β_i. An ON (busy) state represents the period used by primary users and an OFF (idle) state represents the unused period [6], [13].

4.2 Cognitive Radio Capacity Model

In the CR network, the available spectrum bands are not contiguous and may be distributed over a wide frequency range with a different bandwidth. For more flexible manipulation of heterogenous spectrum bands, we employ an orthogonal frequency division multiplexing (OFDM) as a physical layer technology, where each spectrum band i has a different bandwidth B_i, consisting of multiple subcarriers. Usually, each subcarrier has a different channel gain and a noise level that are time-varying. However, when we consider long-term spectrum characteristics, both fast and frequency selective fading effects are mitigated, and hence we can say the channel gain and noise level in the same spectrum are identical over a long-term period. If transmission power is also identical within the spectrum, a normalized channel capacity $c_i(k)$ (bits/sec/Hz) of spectrum band i can be expressed as $c_i(k) = r_i(k)/B_i$, where $r_i(k)$ is the capacity of user k when all subcarriers in spectrum i are assigned to user k.

However, in CR networks, each spectrum i cannot provide its original capacity $c_i(k)$. First, CR users cannot have a reliable spectrum permanently and need to move from one spectrum to another according to the PU activity, which introduces the so-called *spectrum switching delay*. During the switching time, the transmission of the CR user is temporarily disconnected. Here, spectrum switching delay includes times for the spectrum decision process in the base-station, signaling for establishing new channels, and RF front-end reconfiguration. In IEEE 802.22 Wireless Regional Area Network (WRAN), switching delay is required to be less than 2 sec [11]. Also conventional mobile broadcasting systems, for example, Qualcomm's MediaFLO, show an average physical layer channel switching delay up to 1.5 sec [18]. Depending on the development of the hardware technology, we believe that it will be much

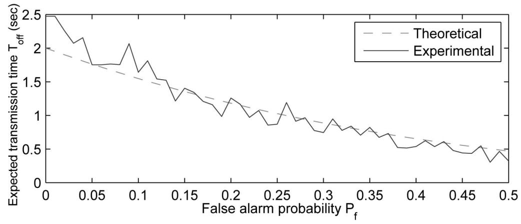


Fig. 3. Expected transmission time in imperfect sensing.

shorter but still be a significant factor to influence the network performance. Furthermore, CR users are not allowed to transmit during sensing operations, leading to the periodic transmissions with sensing efficiency η_i [13].

These unique features in CR networks show a significant influence on the spectrum capacity $\mathbf{C_i(k)}$. To describe all these stochastic activities, we define a new capacity notion, the so-called *CR capacity* $C_i^{\mathrm{CR}}(k)$, which is defined as the expected normalized capacity of user k in spectrum i as follows:

$$C_i^{\mathrm{CR}}(k) = E[\mathbf{C_i(k)}] = \frac{T_i^{\mathrm{off}}}{T_i^{\mathrm{off}} + \tau} \cdot \eta_i \cdot c_i(k), \tag{1}$$

where τ represents the spectrum switching delay, and T_i^{off} is the expected transmission time without switching in spectrum i. Since CR users face to the spectrum switching after the idle period, the first term in (1) represents the transmission efficiency when CR users occupy spectrum i.

If we consider perfect sensing, i.e., both false alarm and detection error probabilities are zero, T_i^{off} is obtained as $1/\beta_i$, which is the average idle period based on the ON-OFF model in Section 3. On the contrary, in case of imperfect sensing, we should account for the influence of sensing capability. Let Δt be a sensing period. Then, the average number of sensing slots in the idle period n_{s} is $\lceil 1/\beta_i/\Delta t \rceil$. From this, the expected transmission time can be obtained as follows:

$$\begin{aligned} T_i^{\mathrm{off}} &= \Delta t \cdot \sum_{k=1}^{n_{\mathrm{s}}-1} k \cdot \left(1 - P_i^{\mathrm{f}}\right)^k \cdot P_i^{\mathrm{f}} + \frac{1}{\beta_i} \cdot \left(1 - P_i^{\mathrm{f}}\right)^{n_{\mathrm{s}}} \\ &= \Delta t \cdot \left[\frac{\left(1 - P_i^{\mathrm{f}}\right)\left(1 - \left(1 - P_i^{\mathrm{f}}\right)^{n_{\mathrm{s}}-1}\right)}{P_i^{\mathrm{f}}} \right. \\ &\qquad \left. - (n_{\mathrm{s}} - 1) \cdot \left(1 - P_i^{\mathrm{f}}\right)^{n_{\mathrm{s}}} \right] + \frac{1}{\beta_i}\left(1 - P_i^{\mathrm{f}}\right)^{n_{\mathrm{s}}}, \end{aligned} \tag{2}$$

where P_i^{f} represents a false alarm probability of spectrum i at each sensing slot. Here, T_i^{off} can be expressed as the sum of the expected durations until when the false alarm is first detected in each slot. As P_i^{f} increases, T_i^{off} decreases, resulting in decrease in CR capacity, which is described in Fig. 3. Here, we consider a cooperative sensing scheme based on "OR" fusion, where its detection error probability converges to 0 as the number of users increases [15]. Thus, the detection error probability can be ignored in estimating CR capacity.

5 Spectrum Decision for Real-Time Applications

Real-time applications are sensitive to delay and jitter. Moreover, they require a reliable channel to support a sustainable rate during the session time. Thus, real-time

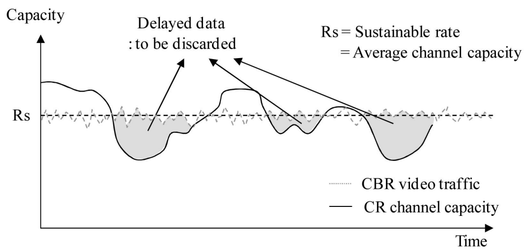


Fig. 4. Data loss in real-time video applications.

applications have strict constraints on the delay bound and the sustainable rate. Generally, real-time applications drop the packets not arrived within the delay bound. Even though the network can support sustainable rate R_s on average, packets can be delayed and finally discarded in the receiver due to the variation of channel capacity, as shown in Fig. 4.

Unlike conventional wireless networks, the CR network has unique delay factors. When CR users either sense or switch the spectrum, they need to stop transmission temporarily, which prevents the real-time application from maintaining its sustainable rate, leading to delay and jitter. To observe the effect of the delay uniquely shown in CR networks, we assume that a buffering scheme is optimized to absorb delay factors in conventional wireless networks, such as application layer, link layer, and transmission delays. Then, the additional delay factors uniquely introduced by CR networks can directly lead to data losses. For this reason, we use the data loss rate to evaluate the service quality of real-time applications. Also real-time applications are assumed to have a set of discrete sustainable rates and to adjust their rates through the negotiation flexibly.

According to the decision events, as explained in Section 3.3, the proposed spectrum decision for real-time application can be classified into an MVSD-SS and an MVSD-MS.

5.1 Minimum Variance-Based Spectrum Decision—Single Selection (MVSD-SS)

Real-time applications need to have more reliable and time-invariant communication channels to satisfy strict service requirements, such as delay constraints and sustainable rates. However, how to maximize the total network capacity is still a crucial problem. To address these issues together, it is essential to guarantee the service quality of real-time applications with minimum spectrum resources. Thus, the spectrum decision problem can be formulated as an optimization to minimize bandwidth utilization subject to the constraints on the sustainable rate, data loss rate, and number of transceivers. However, this problem is mixed with the discrete optimization for spectrum selection and the continuous optimization for bandwidth allocation, which is difficult to solve. Instead, we introduce a three-stage spectrum decision method as follows:

5.1.1 Step 1: Spectrum Selection

From the view of the date loss rate caused by delay, the network prefers the spectrum bands with a lower PU activity. On the other hand, for network capacity, the channel quality needs to be considered in spectrum decision. Thus, to maintain service quality and achieve the maximum network capacity, CR user k selects spectrum bands according to the following linear integer optimization:

$$Maximize\colon \sum_{i\in\mathcal{A}} \frac{C_i^{\mathrm{CR}}(k)}{\beta_i} x_i, \tag{3}$$

$$subject\ to\colon \sum_{i\in\mathcal{A}} x_i = N, \tag{4}$$

$$C_i^{\mathrm{CR}}(k)\cdot W_i \cdot x_i \geq \frac{R_{\mathrm{s}}(k)}{N} \quad (\forall i \in \mathcal{A}), \tag{5}$$

where N is the number of the transceivers of the CR user, and W_i is the currently available bandwidth of spectrum i that is equal or less than the total bandwidth B_i, $\mathcal{A}$ is the set of currently available spectrum bands, and $x_i \in \{0,1\}$ represents the spectrum selection parameter that equals 1 if spectrum i is selected in the binary integer optimization.

This optimization considers both PU activity β_i and CR capacity $C_i^{\mathrm{CR}}(k)$ simultaneously as shown in (3). The number of the selected bands is restricted to the number of transceivers N as given in (4). The last constraint on sustainable rate $R_{\mathrm{s}}(k)$ in (5) ensures that the selected spectrum bands have enough bandwidth for resource allocation, which is explained in Step 2 (Section 5.1.2). Since real-time applications usually require much stricter service requirements than best-effort applications, they have a higher priority for resource allocation. Thus, available bandwidth W_i includes the portions currently occupied by best-effort applications as well as the unused portion of the spectrum.

5.1.2 Step 2: Resource Allocation

Here, the CR network determines the bandwidth, i.e., a set of subcarriers, of the selected spectrum bands to meet the constraints on both sustainable rate $R_{\mathrm{s}}(k)$ and target data loss rate $P_{\mathrm{loss}}^{\mathrm{th}}$. To allocate the bandwidth properly, first, we derive total capacity $\mathbf{R_T}(\mathbf{k})$ and data loss rate $P_{\mathrm{loss}}(k)$ of user k. When bandwidth $w_i(k)$ is allocated to the selected spectrum i for user k, the expected total capacity can be obtained as follows:

$$E[\mathbf{R_T}(\mathbf{k})] = \sum_{i\in\mathcal{S}} C_i^{\mathrm{CR}}(k)\cdot w_i(k), \tag{6}$$

where $\mathcal{S}$ is the set of the selected spectrum bands. To satisfy the service requirement on the sustainable rate, $E[\mathbf{R_T}(\mathbf{k})]$ should be equal to $R_{\mathrm{s}}(k)$.

Unlike total capacity, data loss rate $P_{\mathrm{loss}}(k)$, is expressed in a complicated form, as derived in Appendix A. Thus, it cannot be directly used for the optimization. However, since the variance of the total capacity is proportional to the data loss rate, as shown in Appendix B, we can use the following variance for resource allocation, instead of the data loss rate.

$$\mathrm{Var}[\mathbf{R_T}(\mathbf{k})] = \sum_{i\in\mathcal{S}} \frac{T_i^{\mathrm{off}}\eta_i \cdot \left(T_i^{\mathrm{off}} + \tau - T_i^{\mathrm{off}}\eta_i\right)}{\left(T_i^{\mathrm{off}}+\tau\right)^2}\cdot c_i(k)^2 \cdot w_i(k)^2. \tag{7}$$

Based on the capacity variance obtained above, the CR network determines optimal bandwidth $w_i(k)$ of the selected bands to minimize the variance of the total capacity as follows:

$$Minimize\colon \mathrm{Var}[\mathbf{R_T}(\mathbf{k})], \tag{8}$$

$$\textit{subject to:} \sum_{i=1}^{M} C_i^{\mathrm{CR}}(k) \cdot w_i(k) = R_{\mathrm{s}}(k), \tag{9}$$

$$w_i(k) < W_i \quad (\forall i \in \mathcal{S}). \tag{10}$$

Equations (9) and (10) represent the constraints on the sustainable rate and the available bandwidth, respectively. By the Lagrange multiplier method, optimal bandwidth $w_i(k)$ can be obtained as follows:

$$w_i(k) = \frac{R_{\mathrm{s}}(k) \cdot \left(T_i^{\mathrm{off}} + \tau\right)}{c_i(k) \cdot \eta_i \left(T_i^{\mathrm{off}} + \tau - T_i^{\mathrm{off}} \eta_i\right) \cdot \sum_{i \in \mathcal{S}} \frac{T_i^{\mathrm{off}}}{T_i^{\mathrm{off}} + \tau - T_i^{\mathrm{off}} \eta_i}}. \tag{11}$$

5.1.3 Step 3: QoS Checkup

This optimization is based on the minimum variance, which guarantees the minimum data loss rate but may not satisfy target loss rate $P_{\mathrm{loss}}^{\mathrm{th}}$. If the expected data loss rate $P_{\mathrm{loss}}(k)$ given in (29) is still higher than $P_{\mathrm{loss}}^{\mathrm{th}}$ after this optimization, we need to perform one of the following approaches to satisfy the target loss rate:

- *Aggressive approach:* By sacrificing the bandwidth efficiency, the CR network tries to find the proper spectrum bands to meet the service requirements. To this end, the selected band having the highest PU activity needs to be replaced by the one with the highest $C_i^{\mathrm{CR}}(k)/\beta_i$ among the unselected bands that have a lower PU activity than the original one. If CR network cannot find the proper spectrum band in the aggressive approach, it switches to the conservative approach as explained below.
- *Conservative approach:* Here, real-time applications are assumed to support multiple sustainable rates and to adjust their rates adaptively. Thus, in this approach, instead of increasing the bandwidth, the CR network reduces the current sustainable rate to a one-step lower rate through the renegotiation of the service quality and repeats the MVSD-SS while maintaining the bandwidth efficiency.

Both aggressive and conservative approaches are applied in spectrum decision combining with resource management, which is explained in Section 7.

5.2 Minimum Variance-Based Spectrum Decision—Multiple Selections (MVSD-MS)

MVSD-MS is performed when CR users lose one of their spectrum bands due to either PU activity or quality degradation on that band. Since multiple users need new spectrum bands at the same time, first, they should determine the order of the spectrum decision. Let $R_{\mathrm{lost}}(k)$ be the lost capacity of user k resulting from spectrum switching. Then, the loss rate of user k is obtained as $R_{\mathrm{lost}}(k)/R_{\mathrm{s}}(k)$. In MVSD-MS, CR users are selected in order from the highest to the lowest loss rate. After that, they select a single spectrum with the highest $C_i^{\mathrm{CR}}(k)/\beta_i$ to meet the sustainable rate, and accordingly allocate the bandwidth of all assigned spectrum bands.

6 Spectrum Decision for Best-Effort Applications

The objective of typical scheduling methods for best-effort applications is to maximize the network capacity. The spectrum decision for best-effort applications has the same objective, but additionally needs to exploit the PU activity and long-term channel characteristics. Similar to MVSD in Section 5, spectrum decision for the best-effort application can be classified into a maximum capacity-based spectrum decision—single selection and a multiple selections.

6.1 Maximum Capacity-Based Spectrum Decision—Single Selection (MCSD-SS)

Optimally, for the maximum capacity, the CR network has to perform the spectrum decision over all current transmissions at every decision event, which requires a high computational complexity. Also, the entire resource reallocation leads to the spectrum switching of the multiple users at the same time, resulting in the abrupt quality degradation. Instead, we introduce a suboptimal method for best-effort applications. If current resource allocation is optimal, the spectrum decision to maximize the network capacity can be simplified as the following selection problem to choose spectrum bands so that the decision gain can be maximized.

$$\textit{Maximize:} \sum_{i \in \mathcal{A}} \left(\mathcal{G}\left(i, C_i^{\mathrm{CR}}(k), W_i\right) - \mathcal{L}\left(i, C_i^{\mathrm{CR}}(k), W_i\right)\right) x_i, \tag{12}$$

$$\textit{subject to:} \sum_{i \in \mathcal{A}} x_i = N, \tag{13}$$

where $\mathcal{G}(i, C_i^{\mathrm{CR}}(k), W_i)$ is the expected capacity gain when new user k with CR capacity $C_i^{\mathrm{CR}}(k)$ joins spectrum i with available bandwidth W_i and $\mathcal{L}(i, C_i^{\mathrm{CR}}(k), W_i)$ is the expected capacity loss of other users in that spectrum band. $\mathcal{A}$ is the set of currently available spectrum bands and N is the number of the transceivers of a CR user. $x_i \in \{0, 1\}$ represents the spectrum selection parameter. The decision gain can be defined as the sum of the difference between capacity gain and capacity loss caused by the addition of a new user.

Assume that a spectrum sharing algorithm assigns the bandwidth to the users fairly. Then, the capacity of each user competing for the same spectrum can be approximated as $C_i^{\mathrm{CR}}(k) \cdot W_i / n_{\mathrm{b},i}$ where $n_{\mathrm{b},i}$ represents the number of best-effort users currently residing in spectrum i. Based on this capacity, the decision gain can be derived as follows:

$$\mathcal{G} - \mathcal{L} = \frac{C_i^{\mathrm{CR}}(k) \cdot W_i}{n_{\mathrm{b},i} + 1} - \sum_{j \in \mathcal{E}_i} \left(\frac{1}{n_{\mathrm{b},i}} - \frac{1}{n_{\mathrm{b},i} + 1}\right) \cdot C_i^{\mathrm{CR}}(j) \cdot W_i, \tag{14}$$

where $\mathcal{E}_i$ is the set of the best-effort CR users currently residing in spectrum band i. The first term represents the capacity gain of new CR user k and the second term describes the total capacity loss of other CR users in spectrum i.

6.2 Maximum Capacity-Based Spectrum Decision—Multiple Selections (MCSD-MS)

Similar to the MVSD-MS, MCSD-MS enables multiple CR users to select a single spectrum band. Thus, the CR network first determines the order of the spectrum decision, and then chooses a spectrum band for each CR user as follows:

- Each CR user who loses its spectrum band, finds a candidate spectrum band with the highest decision gain.

- A CR user with the highest decision gain is assigned to the spectrum first through the optimization in (12).
- According to the optimization result, the CR network updates the current bandwidth allocation and repeats the MCSD-MS for the remaining CR users who need to be assigned to a new spectrum band.

7 Dynamic Resource Management for Spectrum Decision

Because of the PU activities, available spectrum bands show time-varying characteristics in the CR network. Thus, with the only proposed decision schemes, the CR network is not able to exploit spectrum resources efficiently, and hence results in the violation of the guaranteed service quality. As a result, the CR network necessitates an additional resource management scheme to coordinate the proposed spectrum decision methods adaptively with bandwidth fluctuations. The main objectives of the proposed resource management are as follows:

- The CR network is capable of determining the acceptance of a new incoming CR user without any effect on the service quality of currently transmitting users.
- During the transmission, the CR network needs to maintain the service quality of currently transmitting users by considering the fluctuation of the available bandwidth.
- Since real-time users usually have a higher priority in spectrum access, best-effort users may not have enough resources. Thus, the CR network may be required to balance the bandwidth between both applications.

In the following sections, we define the network states to describe the current spectrum utilization. Based on these states, we present an admission control scheme, and then propose decision control methods for two different events: CR user and primary user appearances.

7.1 Spectrum States for Resource Management

To exploit spectrum resources efficiently, the proposed spectrum decision needs to adapt to the time-varying network conditions. Thus, we classify the network condition into three states according to the bandwidth utilization. Let W_{R} be the bandwidth currently assigned to real-time users, and W_{av} be the total available bandwidth not occupied by primary users. $W_{\min}$ represents the minimum bandwidth to guarantee the service requirements of current users. W_{R}, W_{av}, and $W_{\min}$ are time-varying according to the spectrum decision results and PU activities. Since best-effort users do not have strict service requirements, we consider only the bandwidth assigned to real-time users in determining the network state. As shown in Fig. 5, the network states are classified as follows:

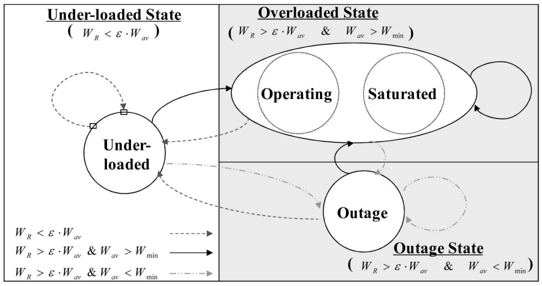


Fig. 5. The state diagram for resource management.

- *Underloaded state:* If the current occupancy of real-time users, $W_{\mathrm{R}}/W_{\mathrm{av}}$ is less than ϵ, the CR network is underloaded. ϵ is the predefined overload threshold to determine if the network is overloaded or not.
- *Overloaded state:* When $W_{\mathrm{R}}/W_{\mathrm{av}} > \epsilon$, the CR network is now overloaded. According to the amount of the remaining bandwidth, this state can be classified into two substates. If the expected bandwidth required for the spectrum decision, W_{req}, is less than the currently unused bandwidth $W_{\mathrm{av}} - W_{\mathrm{R}}$, the CR network is in the beginning of the overloaded state and still has enough resources (*operating state*). Otherwise, the CR network is almost saturated and does not have enough bandwidth for the current spectrum decision operation (*saturated state*). W_{req} is given in Section 7.3.2.
- *Outage state:* If available bandwidth W_{av} is below $W_{\min}$, the CR network cannot provide the guaranteed service quality to the currently active CR users.

If ϵ becomes higher, real-time users can have more stable sustainable rate due to less admission and rate controls, but the outage probability will be higher.

7.2 Admission Control

The CR network is responsible for guaranteeing the service requirements of current CR users regardless of bandwidth fluctuations. Thus, if the CR network cannot maintain the service requirements, it should reject a new incoming CR user, referred to as an *admission control*. The proposed admission control method requires the following procedures:

- *User characterization:* According to the radio condition, each CR user requires different bandwidth to achieve the same service requirements. The radio condition of each user k can be represented as its normalized capacity over all spectrum bands $C(k)$ as follows:

$$C(k) = \frac{\sum_{i=1}^{M} C_i^{\mathrm{CR}}(k) \cdot B_i}{\sum_{i=1}^{M} B_i}, \tag{15}$$

  where M is the number of all spectrum bands and B_i is the total bandwidth of spectrum i.
- *Bandwidth for guaranteeing the service quality:* Since available bandwidth W_{av} varies over time, the CR network cannot always satisfy the service requirements. Thus, we introduce a lower limit of bandwidth $W_{\min}$ to guarantee the service requirements of current CR users. Assume that regardless of the bandwidth fluctuation, the CR network should guarantee an average sustainable rate, $R_{\min}(k)$, over an entire session of real-time user k. Then, the minimum bandwidth of user k to support $R_{\min}(k)$ is expressed as $R_{\min}(k)/C(k)$. When a new CR user appears, $W_{\min}$ can be expressed as the sum of the

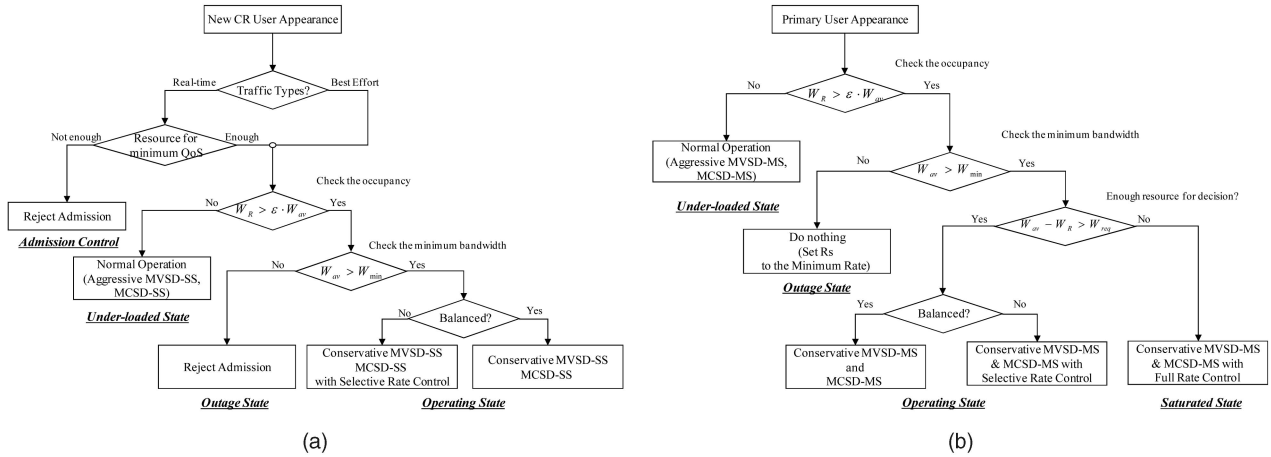


Fig. 6. The flowchart for the proposed decision control: (a) CR user appearance, and (b) primary user appearance.

minimum bandwidths for all CR users including both current and incoming users.

- *Admission criterion:* The proposed spectrum decision is designed for the states when W_{av} is above $W_{\min}$. Otherwise, the network is in the outage state, and hence cannot maintain service requirements of current CR users. However, $W_{\min}$ is time-varying according to the current users and spectrum availability. To mitigate this temporal resource fluctuation, we first determine stable interval $T_{\min}$, which is defined as the average period where no CR user appears, and accordingly $W_{\min}$ does not change. Assume that the departure rate of CR users is μ. Then, $T_{\min}$ can be obtained as $1/(\mu \cdot n_{\mathrm{r}})$ on average, where n_{r} is the number of current real-time users. To avoid resource outage of current CR users, the proposed scheme accepts a new incoming user only if a *resource outage probability* during this interval is greater than the predetermined acceptable outage probability π. Otherwise, it is rejected. The resource outage probability, P_{out}, is the probability that $W_{\mathrm{av}} < W_{\min}$, which is derived in Appendix C.

The performance of the admission control method depends on acceptable outage probability π. If the CR network has a higher π, it can accept more users, resulting in higher quality degradation since it is highly probable that $W_{\mathrm{av}} < W_{\min}$.

The proposed MVSD method explained in Section 5 tries to satisfy target data loss rate $P_{\mathrm{loss}}^{\mathrm{th}}$ on the assumption that the network has sufficient available bandwidth. Thus, the spectrum decision also needs to consider the additional data loss factor resulting from the bandwidth shortage. Since the network capacity is proportional to the available bandwidth, the data loss rate newly introduced by the admission control can be approximately estimated as follows:

$$\widehat{P}_{\mathrm{loss}} = \frac{W_{\min} - E[\mathbf{W}_{\mathrm{av}} | \mathbf{W}_{\mathrm{av}} < W_{\min}]}{W_{\min}} \cdot P_{\mathrm{out}}. \tag{16}$$

Here, the first term represents the ratio of the amount of bandwidth shortage to the bandwidth limit $W_{\min}$.

Then, the actual data loss rate should be expressed as the sum of $P_{\mathrm{loss}}^{\mathrm{th}}$ and $\widehat{P}_{\mathrm{loss}}$. Assume that real-time users have maximum allowable data loss rate P_{loss}. To satisfy this service requirement, target rate $P_{\mathrm{loss}}^{\mathrm{th}}$ should be decided as follows:

$$P_{\mathrm{loss}}^{\mathrm{th}} = P_{\mathrm{loss}} - \widehat{P}_{\mathrm{loss}}. \tag{17}$$

The proposed admission control method is originally designed only for real-time users. Since the best-effort users do not have strict service requirements, they do not need the admission control scheme.

7.3 Decision Control

Here, we propose decision control schemes for both CR and primary user appearances, which enable spectrum decision to adapt to the different network states.

7.3.1 Decision Control in CR User Appearance

One of the important roles in the decision control is how to allocate spectrum resources with the minimum influence on current CR users when a new CR user appears. Fig. 6a shows the procedures of the proposed decision control when the new CR user appears. According to the state, the proposed control scheme coordinates the spectrum decision as follows:

Underloaded state. Since the available bandwidth is sufficient in the underloaded state, the CR network performs the spectrum decision aggressively, i.e., the aggressive MVSD-SS for real-time users and the MCSD-SS for best-effort users.

Overloaded state. Since the available bandwidth becomes scarce in this state, the spectrum decision needs to be more spectrum-efficient. Thus, the CR network performs the conservative MVSD-SS for the real-time user.

However, since real-time users occupy much higher bandwidth through this operation, best-effort users may experience bandwidth starvation in the overloaded state. If the CR network is required to balance the bandwidth between real-time and best-effort users, it needs to check the current bandwidth utilization of both applications before MCSD-SS. Let δ be a balance coefficient predetermined by the CR network. If the average bandwidth of current real-time users, $W_{\mathrm{R}}/n_{\mathrm{r}}$ is greater than the weighted average bandwidth for best-effort users, $\delta \cdot (W_{\mathrm{av}} - W_{\mathrm{R}})/n_{\mathrm{b}}$, current resource allocation is considered to be unbalanced where n_{r} and n_{b} are the numbers of the current real-time and the current best-effort users, respectively. If δ is greater than 1, real-time users can occupy more bandwidth, and hence guarantee more stable service quality.

To solve the resource starvation problem in best-effort users, we propose a *selective rate control* that maintains resource balance in the overloaded state by reducing the sustainable rate of the selected real-time users. When each real-time user k reduces its sustainable rate to a one-step lower rate, the expected bandwidth gain is expressed as $\Delta R(k)/C(k)$ where $\Delta R(k)$ and $C(k)$ is the rate decrement and the normalized capacity of a real-time user k, respectively. Based on the bandwidth gain, the CR network selects real-time users for the selective rate control to minimize total rate reduction subject to the balance constraint, which can be expressed as the following linear integer optimization problem:

$$Minimize\colon \sum_{k\in\mathcal{R}} \Delta R(k)\cdot x_k, \tag{18}$$

$$subject\ to\colon \frac{W_{\mathrm{R}}-\Delta W}{n_{\mathrm{r}}} - \delta\cdot\frac{W_{\mathrm{av}}-W_{\mathrm{R}}+\Delta W}{n_{\mathrm{b}}} \le 0, \tag{19}$$

$$\Delta W \ge \sum_{k\in\mathcal{R}} \frac{\Delta R(k)}{C(k)}\cdot x_k, \quad x_k\in\{0,1\}, \tag{20}$$

where $\mathcal{R}$ is the set of real-time users currently active and ΔW is the bandwidth required for the balance.

The real-time users selected through the above optimization reduce their sustainable rates to the one-step lower rates and then perform the resource allocation explained in Section 5.1. Whenever the best-effort user appears in this state, the network tries to satisfy the balanced condition. However, to avoid the abrupt quality degradation of real-time users, a selective rate control can change the sustainable rate of real-time users to only a one-step lower rate.

Outage state. The service requirements of CR users cannot be guaranteed because of resource shortage. Thus, all incoming best-effort users should be rejected in this state to avoid the overall quality degradation. New real-time users in this state are already rejected through the admission control.

7.3.2 Decision Control in Primary User Appearance

Once the CR network accepts the users, it should guarantee their service requirements during the transmission regardless of the bandwidth variation. Fig. 6b shows the decision control procedure in the primary user appearance. According to the network states, the proposed scheme can be performed as follows:

Underloaded state. Similar to the CR user appearance, the CR network performs the spectrum decision aggressively. For a real-time user, the aggressive MVSD-MS is used whereas the MCSD-MS is executed for a best-effort user.

Operating state. In the overloaded state, the decision control starts to coordinate the bandwidth allocation to maintain the service quality. In the primary user appearance, the overloaded state can be divided into two different substates according to the remaining spectrum resources. In the operating state, the CR network is considered to be overloaded but still has enough resources for spectrum decision, i.e., available bandwidth W_{av} is greater than bandwidth required for spectrum decision, W_{req}. The expected bandwidth for MVSD-MS, W_{req} can be derived as follows:

$$W_{\mathrm{req}} = \sum_{k\in\mathcal{R}_l} \frac{R_{\mathrm{lost}}(k)}{\sum_{i\in\mathcal{A}} C_i^{\mathrm{CR}}(k)\cdot W_i}\cdot (W_{\mathrm{av}}-W_{\mathrm{R}}), \tag{21}$$

where W_i is the available bandwidth of spectrum i currently unused by both primary and real-time CR users and $R_{\mathrm{lost}}(k)$ is the lost capacity of user k due to the PU activities. $\mathcal{A}$ is the set of the currently available spectrum bands and $\mathcal{R}_l$ is the set of the real-time users who lose their spectrum bands, respectively. Here, W_{req} is expressed as the sum of the expected bandwidth of user $k\in\mathcal{R}_l$ required to support $R_{\mathrm{lost}}(k)$. The denominator in the summation in (21) represents the total expected capacity of user k over all currently available spectrum bands.

If the bandwidth of both applications is balanced, the CR network performs a conservative MVSD-MS and an MCSD-MS. Otherwise, it needs a selective rate control before the spectrum decision similar to the case in CR user appearance. The only difference is that a selective rate control is just applied to the real-time users losing one of their spectrum bands to minimize the influence on other real-time users.

Saturated state. The other overloaded state in primary user appearance is the saturated state where the remaining available bandwidth is less than the bandwidth required for the spectrum decision. In this case, real-time CR users cannot find new spectrum bands to maintain their current service requirements, which necessitates the renegotiation of their service requirements.

Let all possible sustainable rates for user k be $\{R_{\mathrm{s},1}(k), R_{\mathrm{s},2}(k),\ldots,R_{\mathrm{s},n_k}(k)\}$, where n_k is the number of all possible sustainable rates. Then, the expected bandwidth of each sustainable rate can be obtained as $R_{\mathrm{s},i}(k)/C(k)$ where $C(k)$ is the normalized capacity of user k given in (15). Based on the expected bandwidth gains in renegotiation, we propose a *full rate control* where the sustainable rates of real-time users currently requesting spectrum decision are optimized to satisfy both bandwidth and balance constraints. This optimization problem is expressed as the following linear integer optimization, the so-called lockbox problem [8].

$$Maximize\colon \sum_{k\in\mathcal{R}_l} R_{\mathrm{s}}(k) = \sum_{k\in\mathcal{R}_l}\sum_{i=1}^{n_k} R_{\mathrm{s},n_k}(k)\cdot x_i(k), \tag{22}$$

$$subject\ to\colon \frac{W_{\mathrm{R}}^s+\widehat{W}_{\mathrm{R}}}{n_{\mathrm{r}}} - \delta\cdot\frac{W_{\mathrm{av}}-W_{\mathrm{R}}^s-\widehat{W}_{\mathrm{R}}}{n_{\mathrm{b}}} < 0, \tag{23}$$

$$\widehat{W}_R < W_{\mathrm{av}} - W_{\mathrm{R}}^s, \tag{24}$$

$$\widehat{W}_R = \sum_{k\in\mathcal{R}_l}\sum_{i=1}^{n_k} \frac{R_{\mathrm{s},i}(k)}{C(k)}\cdot x_i(k), \tag{25}$$

$$\sum_{i=1}^{n_k} x_i(k) = 1 \quad x_i(k)\in\{0,1\}, \tag{26}$$

where $\mathcal{R}_l$ and $\widehat{W}_{\mathrm{R}}$ are the set of the real-time users who lose their spectrum bands and their expected bandwidth. W_{R}^s is the bandwidth of the real-time users not affected by the PU activities. Equation (23) is the constraint on the resource balance explained in Section 7.3.1. Equation (24) is the constraint on the available bandwidth required for the spectrum decision.

Outage state. This state cannot provide a guaranteed service quality any longer. Thus, even though the CR network needs the spectrum decision, all CR users who lose their connections reduce their sustainable rate to the minimum and just wait until the network condition becomes better.

8 Performance Evaluation

8.1 Simulation Setup

Here, we simulate an infrastructure-based CR network consisting of one base-station and multiple CR users. Each user is uniformly distributed over the network coverage with the radius of 2 km. The CR network is assumed to operate in 20 licensed spectrum bands consisting of four VHF/UHF TV, four AMPS, four GSM, four CDMA, and four WCDMA bands. The bandwidth of these bands are 6 MHz (TV), 30 kHz (AMPS), 200 kHz (GSM), 1.25 MHz (CDMA), and 5 MHz (WCDMA), respectively. The PU activities of each spectrum band i, α_i and β_i, are randomly selected over $[0, 1]$. The service rate of CR traffic μ is 0.02, and its arrival rate can be determined according to the average number of users. In the simulations, we assume a lognormal fading channel model, where the noise power is -115 dBm, the shadowing deviation is 4, and the path loss coefficient is set to 4 [19]. Transmission power $P_i^k(f)$ is unity over all frequencies.

Through spectrum sensing, the base-station is already aware of the spectrum availability in its coverage. Sensing efficiency η_i, and false alarm probability P_i^{f} are set to 0.9 and 0.99, respectively. These sensing capabilities are assumed to be identical over all spectrum bands. User-based and the band-based quality degradations, explained in Section 3.3, use the same strategies as primary user and CR user appearances, respectively. Thus, we do not consider them in the simulations.

The real-time application is assumed to support five different bitrates: 64, 128, 256, 512 kbps, and 1.2 Mbps. For the resource management, $W_{\min}$ and $R_{\min}$ are set to 10 MHz and 512 Kbps, respectively. The overloaded threshold ϵ is set to 0.5, the balance coefficient δ is 1. The acceptable data loss rate, P_{loss}, and the acceptable outage threshold π are set to 0.05 and 0.03, respectively.

To evaluate the performance of our spectrum decision framework, we introduce three different cases as follows:

- *Case 1:* CR users exploit all functionalities of the entire spectrum decision framework including MVSD, MCSD, and all resource management functions explained in Sections 5, 6, and 7, respectively.
- *Case 2:* CR users perform the proposed spectrum decision framework without the admission control scheme.
- *Case 3:* CR users use only MVSD and MCSD methods (Case 1 without both admission and decision controls).
- *Case 4:* Instead of the optimization schemes in Section 5.1, the proposed MVSD scheme adopts an exhaustive search to determine proper spectrum bands and their bandwidth, which is optimal for real-time users.

Since there are no previous work related to spectrum decision, we compare our decision framework with two straightforward decision criteria as follows:

- *Case 5 Capacity-based decision:* CR users select the spectrum with the highest channel capacity as follows:

$$\begin{aligned} Maximize&: \sum_{i\in\mathcal{A}} C_i^{\mathrm{CR}}(k)\cdot x_i \\ subject\ to&: \sum_{i\in\mathcal{A}} x_i \le N \quad x_i\in\{0,1\} \\ & \sum_{i\in\mathcal{A}} C_i^{\mathrm{CR}}(k)\cdot W_i\cdot x_i \ge \ R_{\mathrm{s}}(k). \end{aligned} \tag{27}$$

  $\mathcal{A}$ is the set of the currently available spectrum bands. The last constraint is applied only to the real-time users.
- *Case 6 PU activity-based decision:* CR users select the spectrum bands with the lowest PU activity. Instead of the objective function in Case 5, Case 6 uses $\sum_{i\in\mathcal{A}} 1/\beta_i \cdot x_i$.

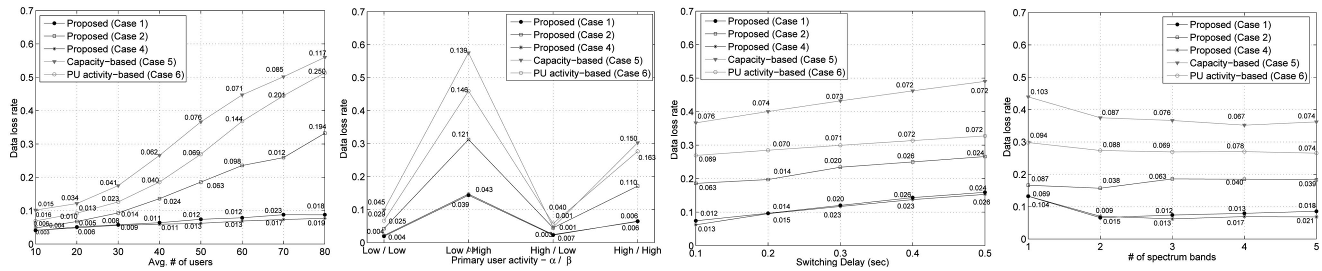
In the following sections, we show the simulation results in three different scenarios (only real-time users, only best-effort users, and both of them).

8.2 Real-Time Applications

First, we consider the scenario with only real-time users to validate the proposed MVSD-SS and MVSD-MS described in Section 5. Since this scenario does not require the decision control for bandwidth balance, Case 3 is not considered in this simulation. The numbers on the graph indicate the standard deviations of each simulation, which show the distribution of the data loss rate over all CR users.

Fig. 7a shows how the average number of users influences the data loss rate. Here, we assume three spectrum bands and 0.1 sec for the switching delay. For this simulation, we generate CR user traffic from 10 to 80 on average. When a small number of users are transmitting, each case shows relatively low data loss rate. However, as the number of users increases, other methods (Cases 2, 5, and 6) increase the data loss rate. On the contrary, Case 1 still maintains a certain level of the data loss rate where the admission control controls the addition of new users adaptively dependent on current network utilization. However, Case 1 shows little higher data loss rate than the acceptable data loss rate. The reason is that during the transmission the MVSD-MS scheme maintains all ongoing transmissions even though they cannot find the spectrum bands to satisfy the acceptable data loss rate, which causes slight increase in the data loss rate. Even though the proposed method does not use admission control (Case 2), it still shows a better data loss rate than Cases 5 and 6.

In Fig. 7b, we investigate the performance of the spectrum decision under four PU activity scenarios—low/low, low/high, high/low, and high/high. Low PU activity (either α_i or β_i) is uniformly distributed between 0 and 0.5, and high PU activity is between 0.5 and 1. The average number of users, the number of spectrum bands, and switching delay are set to 50, 3, and 0.1 sec, respectively. In



(a) (b) (c) (d)

Fig. 7. Data loss rate in real-time applications: (a) average number of users, (b) PU activities, (c) switching delay, and (d) number of spectrum bands.

all cases, Case 1 shows better performance in data loss than other method (Cases 2, 3, 5, and 6), and a similar loss rate to Case 4. Also, is is shown that β_i is a more dominant factor to determine the loss rate than α_i since a higher β_i introduces more frequent switching, leading to a significant performance degradation.

We also show the relationship between the data loss rate and the switching delay in Fig. 7c. Here, we assume 50 users and three spectrum bands. The proposed method (Case 1) shows the lower loss rate than other methods by both rejecting users before the transmission and reducing sustainable rate during the transmission. Case 2 still shows better performance than Cases 5 and 6 because of both MVSD-SS and MVSD-MS. In all cases, a longer switching delay results in a higher data loss rate.

As explained in Section 3.1, the transmission with multiple transceivers can mitigate the effect of capacity fluctuations as well as prevent a temporary disconnection of communication channels. This phenomena are observed in Fig. 7d. Here, we assume 0.1 sec for the switching delay and 50 real-time users. An interesting point is that more spectrum bands do not always lead to good performance in the data loss rate. As the number of spectrum bands increases, the total amount of PU activities over multiple spectrum bands increases, which may cause an adverse effect on the data loss rate. In this simulation, each does not improve its data loss rate significantly when it has more than two spectrum bands.

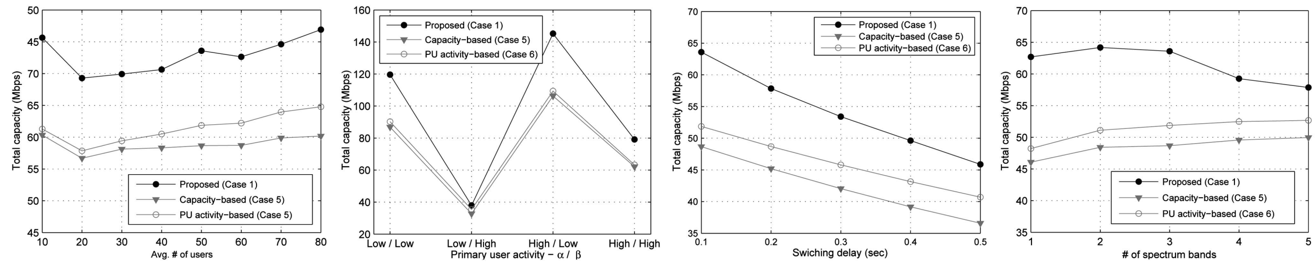
Consequently, in all simulations, the proposed method (Case 1) shows almost the same performance as the optimal method (Case 4), but requires less computational complexity as explained in Section 5.1

8.3 Best-Effort Applications

To evaluate the performance of MCSD-MS and MCSD-SS described in Section 6, we compare the proposed method (Case 1) with Cases 5 and 6. Since the admission and decision control functionalities are not needed in this scenario, we do not consider Cases 2, 3, and 4 here. In this simulation, we also show how the number of users, PU activity, switching delay, and number of spectrum bands influence the total network capacity. As shown in Fig. 8, Case 1 shows higher capacity compared to the capacity-based and PU-activity-based methods. Fig. 8a indicates the relationship between the number of users and total network capacity where Case 1 shows a better performance over Cases 5 and 6 by exploiting the PU activity and the channel condition at the same time. In Fig. 8b, we investigate how PU activities influence the performance of the total capacity. Similarly, Case 1 shows better performance than other cases. Especially, when β_i is lower, Case 1 shows more improvement due to less frequent switching delay. Fig. 8c shows the simulation results on the total network capacity when 50 best-effort users with three spectrum bands are assumed. Here, we observe that increase in switching delay causes an adverse influence on network capacity. Fig. 8d investigates the simulation results on the total network capacity when 50 best-effort users with 0.1 sec switching delay are applied. Similar to the simulation on real-time users, Case 1 shows the best performance in two spectrum bands, but less total capacity in more than two spectrum bands, since it causes a more frequent spectrum switching as well as prevents exploiting channel diversity.

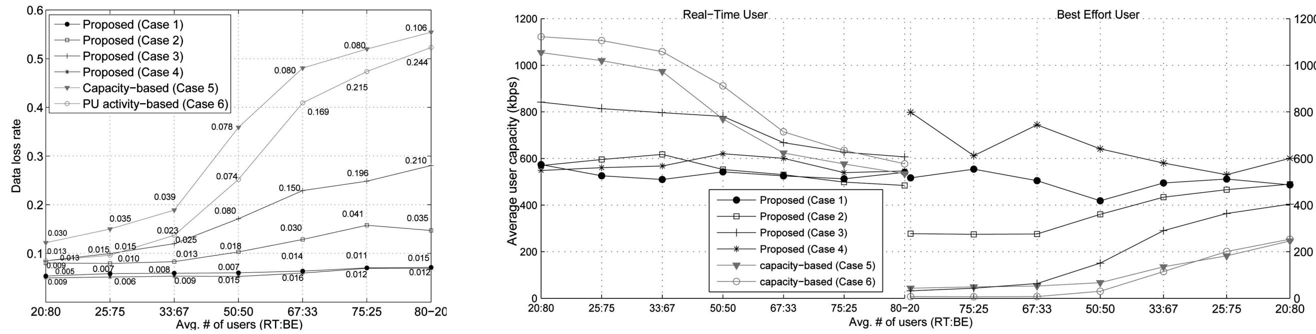
8.4 Hybrid Scenario

Here, we consider a hybrid scenario where both real-time and best-effort users coexist. Similarly, we assume three



(a) (b) (c) (d)

Fig. 8. Total network capacity in best-effort applications: (a) average number of users, (b) PU activities, (c) switching delay, and (d) number of spectrum bands.



(a) (b)

Fig. 9. Performance in the hybrid scenario: (a) data loss rate, and (b) user capacity.

spectrum bands and 0.1 sec switching delay for this simulation. Here, we set the total number of active users to 100 and vary the number of best-effort and real-time users to investigate the performance according to the network state. In Fig. 9a, we show the data loss rate of real-time users on a hybrid scenario. In the underloaded state, i.e., when there less real-time users in the network, we can see each method shows lower data loss rate. On the other hand, overloaded conditions lead to considerably different performance according to the decision methods. Through admission and decision controls, Case 1 admits real-time users only when the network can guarantee the service requirement of current users, and hence maintains the lowest data loss rate. When the proposed method does not use the admission and decision controls (Cases 2 and 3), the CR network accepts much more real-time users than it can provide with the guaranteed service quality, leading to the increase in the data loss rate and the decrease in the average capacity of the real-time user as described in Fig. 9b. On the contrary, Cases 5 and 6 show the worst data loss rates.

In Fig. 9b, we show the average user capacity of real-time and best-effort users in the hybrid scenario. When real-time users are less than best-effort users, Cases 5 and 6 show the highest user capacity in real-time users while maintaining slightly higher data loss rate as that of the proposed methods (Cases 1, 2, and 3). On the contrary, as the number of real-time users increases, Cases 5 and 6 show lower capacity in best-effort users due to the lack of resource management, but the proposed method (Case 1) still provides enough capacity to best-effort users. Even though real-time users occupy most of the bandwidth resources in Cases 5 and 6, they cannot satisfy the service requirements and show the highest data loss rate as observed in Fig. 9a.

By exploiting the admission control scheme, Cases 1 and 2 show better fairness in capacity between both application types while maintaining the low data loss rate in real-time users. Though Case 3 does not use both admission and decision control schemes, it shows slightly higher capacity in best-effort users than Cases 5 and 6 since the MVSD scheme provides bandwidth-efficient resource allocations, leading to increase in available bandwidth for best-effort users as explained in Section 5. Similarly, the optimal method (Case 4) selects the bandwidth-efficient spectrum for real-time users, leading to slightly higher capacity in best-effort users than the proposed method (Case 1), while it achieves almost same data loss rate and user capacity in real-time users as the proposed method.

Fig. 10 shows how the proposed admission control exploits bandwidth resources when 50 real-time users and 50 best-effort users are transmitting simultaneously. From the simulation results, we can see the proposed admission control (Case 1) balances the bandwidth between both applications over the entire simulation time. On the contrary, in Cases 5 and 6, real-time applications occupy most of the available bandwidth to satisfy their service requirements, leading to the bandwidth starvation of best-effort users.

9 Conclusion

In this paper, we introduced a framework for spectrum decision to determine a set of spectrum bands by considering the channel dynamics in the CR network as well as application requirements. To this end, first, a novel spectrum capacity model is proposed that considers unique features in CR networks. Based on this capacity model, an MVSD is developed for real-time applications, which determines the spectrum bands to minimize the capacity variance. For the best-effort applications, an MCSD is proposed where spectrum bands are decided to maximize the total capacity. Moreover, a dynamic resource management scheme is introduced to enable the CR network to

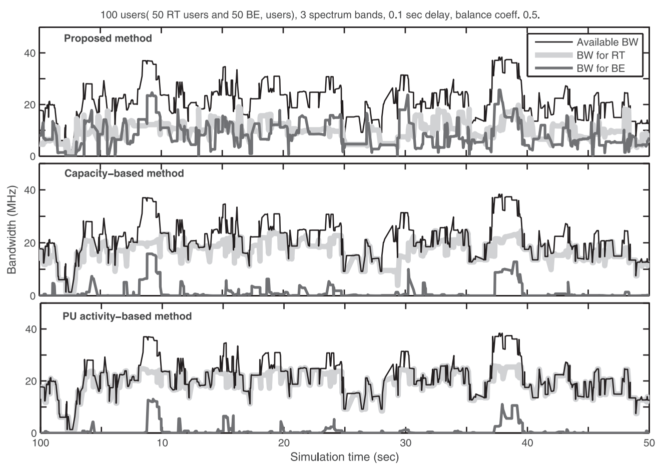


Fig. 10. Bandwidth utilization in the hybrid scenario.

coordinate spectrum decision adaptively dependent on the time-varying spectrum resources. Simulation results show that the proposed spectrum decision framework provides efficient bandwidth utilization while guaranteeing the service quality.

Appendix A

Derivation of the Data Loss Rate in Cognitive Radio Networks

In the CR network, each spectrum band has two discrete capacity states, 0 and $c_i(k) \cdot w_i(k)$ according to its PU activity, as explained in Section 4. Here, $c_i(k)$ and $w_i(k)$ are the normalized capacity and the bandwidth of spectrum i for user k, respectively. Thus, when N spectrum bands are assigned to a CR user k, the total capacity $\mathbf{R_T}(\mathbf{k})$ has 2^N states according to the PU activities of the selected spectrum bands. Thus, each state m has the following state probability:

$$P_m(k) = \prod_{i \in \mathcal{I}_m} \frac{T_i^{\text{off}}}{T_i^{\text{off}} + \tau} \cdot \eta_i \prod_{i \in \mathcal{B}_m} \left(1 - \frac{T_i^{\text{off}}}{T_i^{\text{off}} + \tau} \cdot \eta_i\right), \tag{28}$$

where $\mathcal{I}_m$ and $\mathcal{B}_m$ are the sets of idle spectrum bands and busy spectrum bands at state m, respectively.

Let the sustainable rate of user k be $R_s(k)$ and the capacity of each state m be $\widehat{R}_m(k)$. From the assumption that the data loss occurs when channel capacity is below $R_s(k)$, the data loss rate can be defined as the ratio of the expected capacity loss to the sustainable rate $R_s(k)$ as follows:

$$\begin{aligned} P_{\text{loss}}(k) &= \frac{R_s(k) - \sum_{m=1}^{2^N} \min(R_s(k), \widehat{R}_m(k)) P_m(k)}{R_s(k)} \\ &= \frac{\sum_{m=1}^{2^N} |R_s(k) - \widehat{R}_m(k)| P_m(k)}{2R_s(k)}. \end{aligned} \tag{29}$$

Appendix B

Derivation of the Capacity Variation in Cognitive Radio Networks

From the capacity state probability, derived in (28), the variance of the total capacity $\mathbf{R_T}(\mathbf{k})$ can be derived as follows:

$$\text{Var}[\mathbf{R_T}(\mathbf{k})] = \sum_{m=1}^{2^N} (\widehat{R}_m(k) - R_s(k))^2 \cdot P_m(k). \tag{30}$$

By comparing (29) with (30), we can see that the variance of the total capacity $\text{Var}[\mathbf{R_T}(\mathbf{k})]$ is proportional to the data loss rate $P_{\text{loss}}(k)$. As a result, we can use the capacity variance for resource allocation, instead of the data loss rate. To apply the variance in (30) for the optimization, we need another form of the variance expressed in terms of the bandwidth $w_i(k)$ and the normalized capacity $c_i(k)$ of each spectrum. Since the spectrum is independent with each other, the variance of the total capacity in the selected spectrum bands can be expressed as follows:

$$\begin{aligned} \text{Var}[\mathbf{R_T}(\mathbf{k})] &= \text{Var}\left[\sum_{i \in \mathcal{S}} \mathbf{C_i}(\mathbf{k}) \cdot w_i(k)\right] \\ &= \sum_{i \in \mathcal{S}} \text{Var}[\mathbf{C_i}(\mathbf{k}) \cdot w_i(k)] \\ &= \sum_{i \in \mathcal{S}} (E[(\mathbf{C_i}(\mathbf{k}) \cdot w_i(k))^2] - E[\mathbf{C_i}(\mathbf{k}) \cdot w_i(k)]^2) \\ &= \sum_{i \in \mathcal{S}} \left(\left(c_i(k)^2 \cdot w_i(k)^2 \cdot \frac{T_i^{\text{off}}}{T_i^{\text{off}} + \tau} \cdot \eta_i \right.\right. \\ &\quad \left.\left. - \left(c_i(k) \cdot w_i(k) \cdot \frac{T_i^{\text{off}}}{T_i^{\text{off}} + \tau} \cdot \eta_i\right)^2\right)\right) \\ &= \sum_{i \in \mathcal{S}} \frac{T_i^{\text{off}} \eta_i (T_i^{\text{off}} + \tau - T_i^{\text{off}} \eta_i)}{(T_i^{\text{off}} + \tau)^2} c_i(k)^2 w_i(k)^2, \end{aligned} \tag{31}$$

where $\mathbf{C_i}(\mathbf{k})$ is the random variable to represent the capacity of spectrum i for user k. $\mathcal{S}$ is the set of the selected bands.

Appendix C

Derivation of the Resource Outage Probability

To model PU activities in the spectrum, we can use a two-state Markov chain with the transition probabilities from idle to idle $x_i^{00} = 1 - e^{-\beta_i \Delta t}$, from idle to busy $x_i^{01} = e^{-\beta_i \Delta t}$, from busy to idle $x_i^{10} = e^{-\alpha_i \Delta t}$, and from busy to busy $x_i^{11} = 1 - e^{-\alpha_i \Delta t}$, where Δt is a sensing period. Then, the idle probability of spectrum i after $r\Delta t$, $P_i^{\text{idle}}(r)$, can be expressed as either one of the following probabilities [14]:

$$\begin{aligned} P_i^{\text{i2i}}(r) &= \frac{x_i^{10}}{x_i^{01} + x_i^{10}} + \left(1 - x_i^{01} - x_i^{10}\right)^r \cdot \frac{x^{01}}{x^{01} + x_i^{10}}, \\ P_i^{\text{i2b}}(r) &= \frac{x_i^{10}}{x_i^{01} + x_i^{10}} - \left(1 - x_i^{01} - x_i^{10}\right)^r \cdot \frac{x^{10}}{x^{01} + x_i^{10}}, \end{aligned} \tag{32}$$

where $P_i^{\text{i2i}}(r)$ and $P_i^{\text{i2b}}(r)$ are the expected idle probabilities after $r\Delta t$ when current spectrum states are idle and busy, respectively. If a false alarm probability P_i^{f} is considered, the idle probability of spectrum i can be expressed as either $(1 - P_i^{\text{f}})P_i^{\text{i2i}}(r)$ or $(1 - P_i^{\text{f}})P_i^{\text{i2b}}(r)$.

Based on these probabilities, we derive the expected resource outage probability as follows: Since the network has M spectrum bands, it has 2^M states according to the status of each band. Let $\mathcal{L}$ be a set of states that experience resource outage, i.e., that $W^{\text{av}} < W_{\text{min}}$. $\mathcal{I}_n$ represents a set of idle spectrum bands at state n. Then, resource outage happens when all spectrum bands in $\mathcal{I}_n, n \in \mathcal{L}$ are idle and the rest of bands $i \notin \mathcal{I}_n, n \in \mathcal{L}$ are busy. From this, the resource outage probability after $r\Delta t$, $P_{\text{out}}(r)$ can be derived as follows:

$$P_{\text{out}}(r) = \sum_{n \in \mathcal{L}} \prod_{i \in \mathcal{I}_n} P_i^{\text{idle}}(r) \prod_{i \notin \mathcal{I}_n} \left(1 - P_i^{\text{idle}}(r)\right). \tag{33}$$

Based on this probability, we can obtain the expected resource outage probability during $r\Delta t$, P_{out} as $\sum_{r'=1}^{r} P_{\text{out}}(r')/r$.

Acknowledgments

This material is based upon work supported by the US National Science Foundation under Grant No. CNS-07251580.

REFERENCES

[1] I.F. Akyildiz, W.-Y. Lee, M.C. Vuran, and S. Mohanty, "A Survey on Spectrum Management in Cognitive Radio Networks," *IEEE Comm. Magazine,* vol. 46, no. 4, pp. 40-48, Apr. 2008.

[2] D. Cabric, S.M. Mishra, and R.W. Brodersen, "Implementation Issues in Spectrum Sensing for Cognitive Radios," *Proc. IEEE Asilomar Conf. Signals, Systems and Computers,* pp. 772-776, Nov. 2004.

[3] D. Cabric, S.M. Mishra, D. Willkomm, R. Brodersen, and A. Wolisz, "A Cognitive Radio Approach for Usage of Virtual Unlicensed Spectrum," *Proc. 14th IST Mobile and Wireless Comm. Summit,* June 2005.

[4] L. Cao and H. Zheng, "Distributed Spectrum Allocation via Local Bargaining," *Proc. IEEE Sensor and Ad Hoc Comm. and Networks (SECON),* pp. 475-486, Sept. 2005.

[5] L. Cao and H. Zheng, "Distributed Rule-Regulated Spectrum Sharing," *IEEE J. Selected Areas in Comm.,* vol. 26, no. 1, pp. 130-145, Jan. 2008.

[6] C. Chou, S. Shankar, H. Kim, and K.G. Shin, "What and How Much to Gain by Spectrum Agility?" *IEEE J. Selected Areas in Comm.,* vol. 25, no. 3, pp. 576-588, Apr. 2007.

[7] R. Etkin, A. Parekh, and D. Tse, "Spectrum Sharing for Unlicensed Bands," *IEEE J. Selected Areas in Comm.,* vol. 25, no. 3, pp. 517-528, Apr. 2007.

[8] J.R. Evans and E. Minieka, *Optimization Algorithms for Networks and Graphs,* second ed. CRC Press, 1992.

[9] FCC, ET Docket No 02-135, Spectrum Policy Task Force Report, Nov. 2002.

[10] M. Gandetto and C. Regazzoni, "Spectrum Sensing: A Distributed Approach for Cognitive Terminals," *IEEE J. Selected Areas in Comm.,* vol. 25, no. 3, pp. 546-557, Apr. 2007.

[11] *IEEE P802.22/D0.3.8.1, IEEE 802.22 WG, Draft Standard for Wireless Regional Area Networks Part 22: Cognitive Wireless RAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Policies and Procedures for Operation in the TV Bands,* IEEE, Sept. 2007.

[12] X. Kang, Y. Liang, A. Nallanathan, H. Garg, and R. Zhiang, "Optimal Power Allocation for Fading Channels in CR Networks: Ergodic Capacity and Outage Capacity," *IEEE Trans. Wireless Comm.,* vol. 8, no. 2, pp. 940-950, Feb. 2009.

[13] W.-Y. Lee and I.F. Akyildiz, "Optimal Spectrum Sensing Framework for Cognitive Radio Networks," *IEEE Trans. Wireless Comm.,* vol. 7, no. 10, pp. 3845-3857, Oct. 2008.

[14] W.-Y. Lee and I.F. Akyildiz, "Spectrum-Aware Mobility Management in Cognitive Radio Cellular Networks," to be published.

[15] Y.C. Liang, Y. Zeng, E. Peh, and A.T. Hoang, "Sensing-Throughput Tradeoff for Cognitive Radio Networks," *IEEE Trans. Wireless Comm.,* vol. 7, no. 4, pp. 1326-1337, Apr. 2008.

[16] N. Nie and C. Comaniciu, "Adaptive Channel Allocation Spectrum Etiquette for Cognitive Radio Networks," *Proc. First IEEE Int'l Symp. New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05),* pp. 269-278, Nov. 2005.

[17] C. Peng, H. Zheng, and B.Y. Zhao, "Utilization and Fairness in Spectrum Assignment for Opportunistic Spectrum Access," *ACM Mobile Networks and Applications,* vol. 11, no. 4, pp. 555-576, Aug. 2006.

[18] M.R. Chari, F. Ling, A. Mantravadi, R. Krishnamoorthi, R. Vijayan, G.K. Walker, and R. Chandhok, "FLO Physical Layer: An Overview," *IEEE Trans. Broadcasting,* vol. 53, no. 1, pp. 145-159, Mar. 2007.

[19] T. Rappaport, *Wireless Communications: Principles and Practice,* second ed. Prentice Hall, 2001.

[20] H. Shiang and M. Schaar, "Queuing-Based Dynamic Channel Selection for Heterogeneous Multimedia Applications over Cognitive Radio Networks," *IEEE Trans. Multimedia,* vol. 5, no. 10, pp. 896-909, Aug. 2008.

[21] K. Sriram and W. Whitt, "Characterizing Superposition Arrival Processes in Packet Multiplexers for Voice and Data," *IEEE J. Selected Areas in Comm.,* vol. 4, no. 6, pp. 833-846, Sept. 1986.

[22] L. Zhang, Y. Liang, and Y. Xin, "Joint Beamforming and Power Allocation for Multiple Access Channels in Cognitive Radio Networks," *IEEE J. Selected Areas in Comm.,* vol. 26, no. 1, pp. 38-51, Jan. 2008.

Won-Yeol Lee received the BS and MS degrees from the Department of Electronic Engineering, Yonsei University, Seoul, Korea, in 1997 and 1999, respectively, and the PhD degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, in 2009 under the guidance of Professor Ian F. Akyildiz. From 1999 to 2004, he was the senior research engineer of the Network R&D Center and Wireless Multimedia Service Development Division at LG Telecom, Seoul, Korea. Currently, he is the deputy director of the Wireless Technology Department, Central R&D Laboratory, Korea Telecom (KT), Seoul. His current research interests include cognitive radio networks, next generation wireless systems, and wireless sensor networks. He received the 2008 Researcher of the Year Award in the Broadband Wireless Networking Laboratory, School of Electrical and Computer Engineering, Georgia Institute of Technology. He is a student member of the IEEE.

Ian F. Akyildiz is the Ken Byers Distinguished Chair professor with the School of Electrical and Computer Engineering, Georgia Institute of Technology (Georgia Tech). Since June 2008, he is an honorary professor with the School of Electrical Engineering at the Universitat Politecnico de Catalunya, Barcelona, Spain. Also, since March 2009, he has been an honorary professor with the Department of Electrical, Electronic, and Computer Engineering at the University of Pretoria, South Africa. He is the editor-in-chief of the *Computer Networks (COMNET)* journal as well as the founding editor-in-chief of the *Ad Hoc Networks* journal and the *Physical Communication* journal, all with Elsevier. His current research interests are in cognitive radio networks, wireless sensor networks, and nanocommunication networks. He has received numerous awards, including the 1997 IEEE Leonard G. Abraham Prize Award (IEEE Communications Society) for his paper entitled "Multimedia Group Synchronization Protocols for Integrated Services Architectures" published in the *IEEE Journal on Selected Areas in Communications* in January 1996; the 2002 IEEE Harry M. Goode Memorial Award (IEEE Computer Society) with the citation "for significant and pioneering contributions to advanced architectures and protocols for wireless and satellite networking"; the 2003 IEEE Best Tutorial Award (IEEE Communication Society) for his paper entitled "A Survey on Sensor Networks," published in the *IEEE Communications* magazine in August 2002; the 2003 ACM Sigmobile Outstanding Contribution Award with the citation "for pioneering contributions in the area of mobility and resource management for wireless communication networks"; the 2004 Georgia Tech Faculty Research Author Award for his "outstanding record of publications of papers between 1999-2003"; the 2005 Distinguished Faculty Achievement Award from School of Electrical and Computer Engineering, Georgia Tech; the 2009 Georgia Tech Outstanding Doctoral Thesis Advisor Award for "his 20+ years service and dedication to Georgia Tech and producing outstanding PhD students"; and the 2009 ECE Distinguished Mentor Award from the School of Electrical and Computer Engineering, Georgia Tech. He has been a fellow of the ACM since 1996, and he is a fellow of the IEEE.